AN ALGORITHMIC APPROACH FOR SIMULATING REALISTIC IRREGULAR TILINGS

Por

Alejandro Betancourt

Proyecto de investigación enviado al departamento de

ciencias básicas como requerimiento parcial para la obtención del

título de:

MÁSTER

EN   MATEMÁTICAS APLICADAS

Supervisada y dirigida por:
Juan Carlos Duque y Freddy Hernan Marín

Elaborada en el grupo de investigación:
RiSE (Research in Spatial Economics)

Universidad EAFIT
Medellín, Colombia

Septiembre 2012
(Para graduación en Diciembre 2012)

Este trabajo esta dedicado a mi familia por sus esfuerzos y sacrificios durante mis estudios, sin ellos nada de esto sería posible.

ACKNOWLEDGMENT

ABSTRACT

There exist a wide variety of experiments, or simulations, in which regional scientists require regular and irregular lattice structures with a predefined number of spatial units (n). While many GIS software offer the possibility to generate a regular lattice, to obtain an irregular tiling of a given size is not always a straightforward task. The utilization of an irregular lattice usually requires that the researcher finds a large enough real map that matches as closely as possible the required number of spatial units. This practice is usually carried out without considering whether or not the selected lattice is a good representation, in terms of its connectedness characteristics, of a "typical" real map of polygons.

In this project we propose an algorithm, written in Python as a toolbox for clusterpy, that combines fractal theory, theory of stochastic processes and computational geometry for simulating realistic irregular tiling with a predefined number of spatial units. The irregular tiling generated with this algorithm guarantee consistency in their connectedness characteristics and, therefore, reduce potential distortions in the results of the experiments due to an inappropriate selection of the lattice structure.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

A main task in order to propose regional science methodology is to guarantee a good behaviour of them over real data. This task becomes more difficult when we take into account the fact that the geographies used to test algorithms and methodologies could mislead their valuation in different ways such as the significance of hypothesis tests (Anselin and Moreno, 2003; Mur Lacambra, 1992), the computationally efficiency (Duque et al., 2010), or could violate some of the assumptions used to construct the mathematical theories as the assumption of stationarity in spatial econometrics (Hooper and Hewings, 1981). In order to avoid these consequences different types of tilings such as real maps from specific parts of the world, regular tilings, or Voronoi diagrams, are commonly used by regional scientists however because their topologycal characteristics are far of being considered as those of real maps then the consecuences does not dissapear.

The most intuitive approach is using real maps divided by normative regions; e.g; Europe by countries, United States by counties, Spain by municipalities.[1] In one hand these structures has random realistic connectedness characteristics but in the other hand they do not necessary represent the topological characteristics of different parts of the world, as example, Fig. 1.1 shows two set of 900 areas extracted from two different parts of the world, on the left from the normative division of United States by counties with an average number of neighbors of 5.81 ± 0.96 and maximum of neighbors of 9, on the right from Spain divided by municipalities with an average number of neighbors of 6.015 ± 2.60 and a maximum number of neighbors of 31. These differences are more critical because it is common to find that researchers usually use maps from the countries where they reside. Another disadvantage with this approach is that it could be difficult to construct large sets of maps, even more if different sizes are required. Last problem is

---

[1]According to Duque et al. (2007) normative regions are defined as: *"the expression of a political will; their limits are fixed according to the tasks allocated to the territorial communities, to the sizes of population necessary to carry out these tasks efficiently and economically, or according to historical, cultural and other factors"*

usually faced with some sampling approaches (Anselin and Moreno, 2003).

(a) United States   (b) Spain
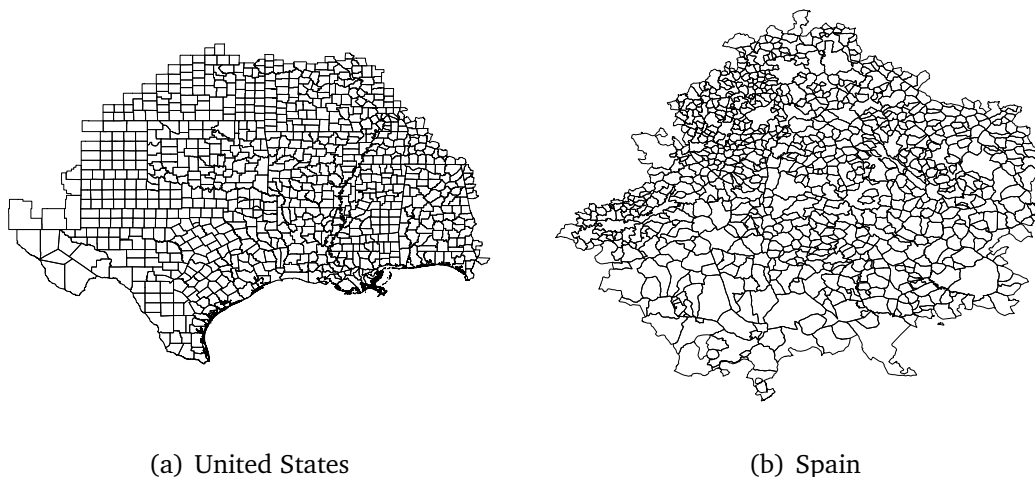
Fig. 1.1: Randomly sampled maps of 900 areas

Regarding regular lattices, they are easy to be generated, do not present restrictions about the number of areas and their over simplified structure usually make it easier to summarize mathematical theories and explain the behaviour of tested methodologies (Whittle, 1954; Bartlett, 1975; Griffith, 1987). The disadvantages of using these tilings are that they make it necessary to set stricter assumptions when the tested methodologies are going to be used for real applications; e.g, the problem of spatial stationarity in STARMA models (Hooper and Hewings, 1981), improper conclusions of the properties of statistics in hypothesis testing (Anselin and Moreno, 2003; Mur Lacambra, 1992) and over qualification of computational efficiency of algorithms (Duque et al., 2010; Aldstadt and Getis, 2006). A middle point between regular tilings and real maps are Voronoi diagrams, which are theoretical structures with random properties that do not restricts the amount and sizes of maps, however as is showed in section 3.2 their topological characteristics are far of those of real maps.[2]

Due to the inappropriateness of using these tilings and the importance of using general realistic structures to evaluate algorithms and methodologies in re-

---

[2]Voronoi diagrams could be considered as an special case of a birth and growth stochastic process known as Johnson-Mehl tessellations when both, birth and growth rates, are equal for each area (Moller, 1995; Chiu, 2011)

gional science, it becomes evident the necessity of a new method to generate tilings with, in average, similar characteristics of real maps from different parts of the world, without restricting the size and number of maps. In order to meet this necessity in this project we propose and algorithmic approach supported in fractal theory, stochastic equations and geometric computation. The algorithm is developed in python and is officially introduced in clusterpy (version 1.1.0)(Duque et al., 2011b).

In chapter 2 of this thesis the basic definitions of polygons and tilings are presented, proposing a consensus taxonomy of tilings, also we propose a set of indicators commonly used to characterize tilings in different fields such as graph theory (Garrison, 1964; Tinkler, 1972), geographical sciences (Gould, 1967; Boots, 1982, 1984, 1985), and metallography (Le Caer and Delannay, 1995; Aste et al., 1996; Peshkin et al., 1991; Le Caer and Delannay, 1993). In chapter 3 the relationship between tilings and regional science is exposed, latter the average values of the indicators of maps from different parts of the world is presented as the benchmark values and compared with the indicators of commonly used tilings. In chapter 4 a brief theoretical background of Stochastic Diferential Equation and Fractal dimension is presented, the rest of the chapter propose two algorithms, the first one (*MR-Polygon*) to generate irregular polygons and the second (*RI-Map*) to generate tilings with similar characteristics of real maps using *MR-Polygons*. In chapter 5 the indicators of *RI-Maps* are obtained and compared with those of regular tilings, Voronoi diagrams and real maps. In chapter 6 and 7 some conclusions and ideas for future research are presented. Finally in chapter 8 is showed how to install the clusterpy and some examples are presented.

2. About tilings and space partitions

2.1   Space partition, consensus taxonomy of tilings

In order to define tessellations, also known as tilings, it is necessary to briefly introduce the concept of polygon, which is a plane figure enclosed by the set of finite straight line segments. Polygons can be categorized according to its boundaries, convexity and symmetry.

i) Respect to its boundary, a polygon can be classified as **simple** if it is formed just by one simple plain figure without holes (Fig. 2.1 a,b and c). Non-simple polygons, also known as **complex**,[3] are all of those that have holes or are formed by multiple parts (Fig 2.1.d). The concept of complex polygons is commonly used in geography to define areas with lakes or islands.

ii) According to its convexity, polygons can be convex or non-convex. Polygons are **convex** if every pair of points in the polygon can be connected by a straight line without crossing the boundary of the polygon (Fig. 2.1 a and b), (Johnson, 2001), otherwise are **non-convex** (Fig. 2.1 c and d).[4]

iii) Based on its symmetry, polygons on which all the angles have the same magnitude and all the sides are of equal length are **regular** (Fig. 2.1.a), otherwise, **irregular** (Fig.2.1 b,c and d) (Johnson, 2001).

A tessellation, in a bidimensional space, is a set of polygons with no gaps and no overlaps covering a subspace or the entire space. Formally, a tessellation of any subspace $S \subseteq R^n$ is the division into k subsets $i \subseteq S$ such that $\cup_i = S$ and $\cap_i = \phi$ where ϕ is the empty set of R^n (Grunbaum and Shephard, 2011), hereinafter only cases where $n = 2$ are taken into account. This definition does not restricts the shapes of the polygons which form the tessellation (Ghyka, 2004; Penrose, 1974; Berger, 1966; Robinson and Berkeley, 1971).[5]

---

[3]Complex polygons do not refer to polygons which exist in the Hilbert plane (Coxeter, 1974).
[4]Concave polygons are simple and non-convex.
[5]Tilley (2006) imposes the condition that a tessellation must be formed by identically shaped polygons

(a) Simple Regular convex (b) Simple Irregular convex

(c) Simple Irregular concave (d) Complex polygon of two parts, one of them with a hole
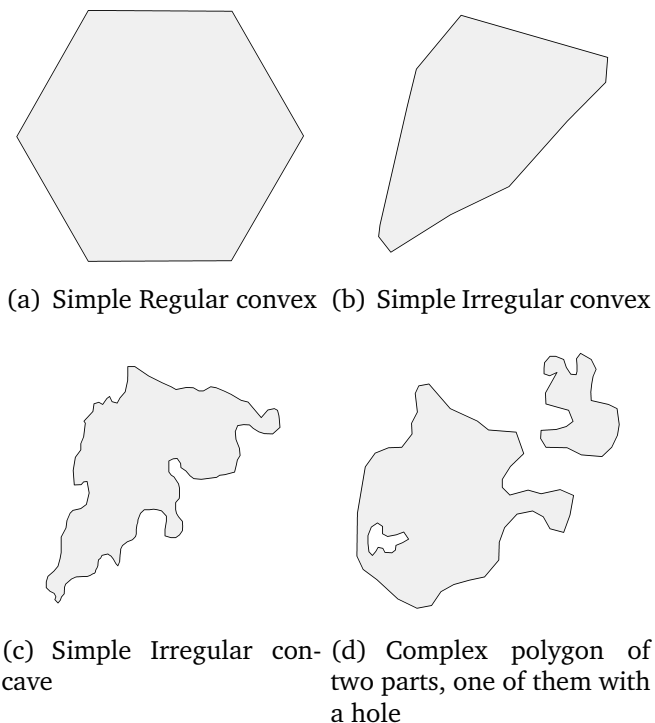
Fig. 2.1: Examples of different type of polygons

There exist different taxonomies of tilings, the use of one or another depends on the study field. In Fig. 2.2 consensus taxonomy is proposed classifying tilings according to the shape of its polygons, the spatial relationship between them, and the use, or not, of symmetric relationships to construct the tessellation.[6] Table 2.1 shows an example of each category of this consensus taxonomy.

i) According to the variety of the shapes of polygons forming the tiling: If all the polygons have the same shape, then the tiling is denoted as **homomorphism**, otherwise **polymorphism**.

ii) According to the regularity of the polygons forming the tiling, and the way that they intersect each vertex, considering a vertex as all the points of tiling that intersects three or more polygons: **Regular** tilings are formed by regular polygons in which all the vertex join the same arrangement of polygons Tilley

---

[6]An alternative category is proposed for tilings formed by fractal polygons that are informally defined by Mandelbrot (1982) as rough fragmented geometric shapes which could be infinitely divided into scalable parts.
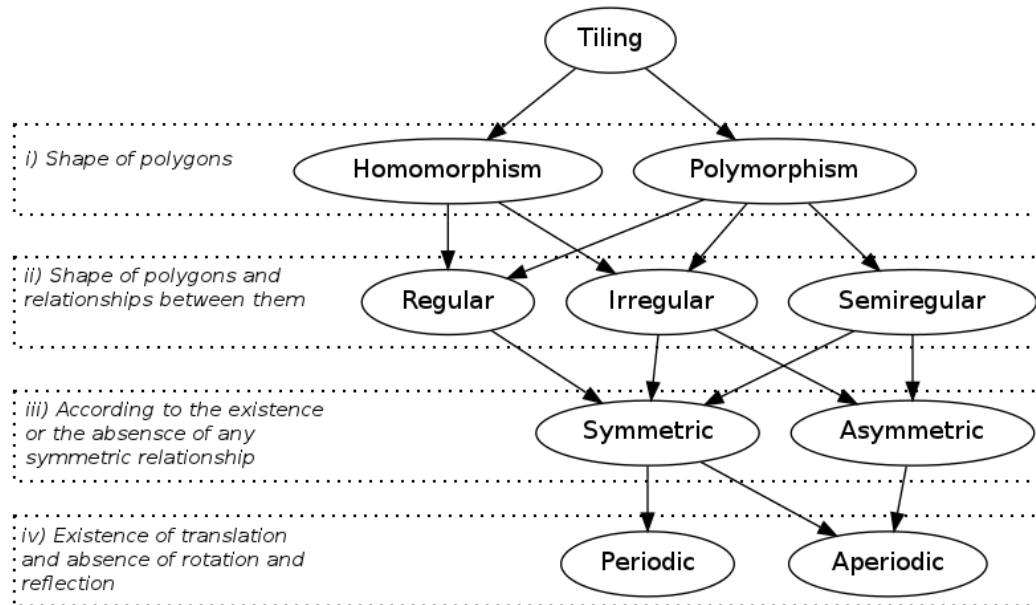
Fig. 2.2: Consensus taxonomy of tilings

(2006). If the polygons are regular, but there are different configurations of vertexes, then the tiling is **semi-regular**, otherwise **irregular** (Ghyka, 2004).

iii) According to the existence of symmetric relationships within the tiling. Symmetrical relationships are three: *Translation:* the tiling is formed by translating a subset of areas; *reflection:* there are axes of reflection in the tiling; *rotation:* it is possible to obtain the same tiling after a rotation process of less than 2π (Radin, 1993). A symmetric tiling implies the presence of, at least, one symmetric relationship, otherwise the tiling is asymmetric.

iv) According to the symmetric relationship of translation. A tiling is **periodic** if and only if it implies the use of translation without rotation or reflection, otherwise **aperiodic** (Tilley, 2006).

2.2 Measuring the topological characteristics of tilings

A common framework in fields such as regional sciences and metallography is to use some sparse matrix indicators to summarize and compare the topological
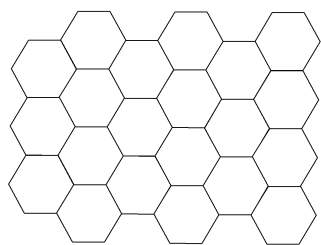
(a) Homorphism
Regular
Periodic Symmetric
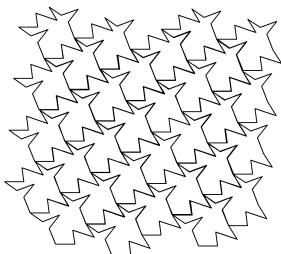
(b) Homorphism
Irregular
Periodic Symmetric
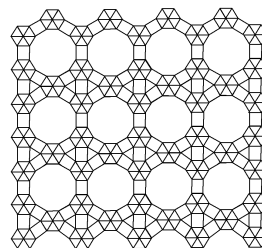
(c) Polymorphism
Semiregular
Periodic Symmetric.
(Ghyka, 2004)
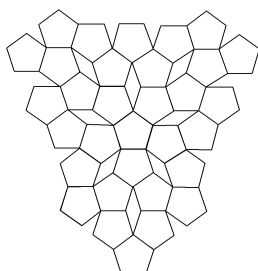
(d) Polymorphism
Semiregular
Aperiodic Symmetric
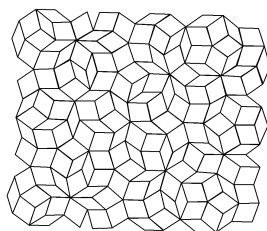
(e) Polymorphism
Semiregular
Aperiodic Asymmetric
(Penrose, 1974)
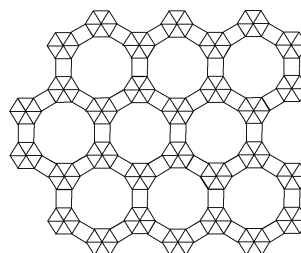
(f) Polymorphism
Regular
Periodic Symmetric
(Ghyka, 2004)
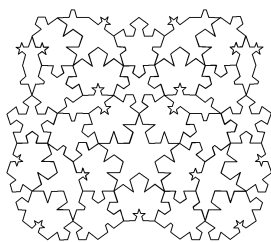
(g) Polymorphism
Irregular
Aperiodic Symmetric
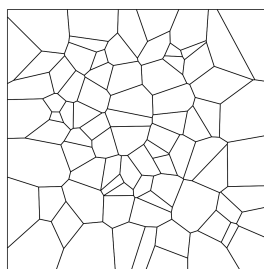
(h) Polymorphism
Irregular
Aperiodic Asymmetric
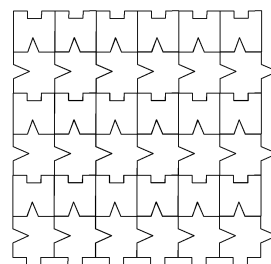
(i) Polymorphism
Irregular
Periodic Symmetric

Table 2.1: Tilings examples

characteristics of tilings through its contiguity matrix (Gould, 1967; Boots, 1982; Le Caer and Delannay, 1995; Aste et al., 1996; Peshkin et al., 1991).[7] In our case six indicators are used, each of which is briefly described below.

The first three indicators are the maximum (M_n), minimum (m_n) and average (μ_1) number of neighbors per area. The fourth indicator, Sparseness S, is the percentage of non-zero elements of the contiguity matrix. The fifth indicator is the first eigenvalue of W (λ_1) which, due to its ability to summarize different aspects of the structures, were originally used to study the connectivity matrix of networks in graphs theory (Garrison, 1964; Tinkler, 1972), and subsequently adopted by regional scientist to characterize topological characteristics of the contiguity matrix (Gould, 1967; Boots, 1982, 1984, 1985). The last indicator is the variance of the number of neighbors per polygon (μ_2) which is commonly used as a measurement of the spatial disorder withing a tiling.

---

[7]For more information about contiguity matrix see Anselin (1988)

3. Tilings in regional science

Within the field of regional science tilings are commonly used with two purposes, in one hand, real maps are used to study real phenomena; e.g. analyze spatial patterns, confirm spatial relationships between variables, detect spatio-temporal regimes within a data panel. On the other hand, they are used to evaluate the behaviour of statistical tests (Anselin and Moreno, 2003; Mur Lacambra, 1992); algorithms (Duque et al., 2010); and topological characteristics of maps (Aste et al., 1996; Le Caer and Delannay, 1995, 1993). In the last case it is necessary to use sets of tilings that satisfy some requirements imposed by the regional scientist; e.g, number of areas, regularity or irregularity of the polygons, number of maps, among others. To accomplish this it is common to use *polymorphism irregular aperiodic asymmetric*, such as real maps and Voronoi diagrams; or *homomorphism regular periodic symmetric* such as regular lattices, as geographical base for real or simulated data. In next section we will be focused on the second use of tilings.

3.1 Commonly used tilings

The process of propound an appropriate set of tilings to evaluate methodologies is not always a trivial task, specially when different topological characteristics may differ even within tilings considered in the same category; e.g; real maps from different parts of the world; regular lattices formed by hexagons or squares. An ideal scenery is to be able to collect a set of tilings with representative characteristics of maps from different parts of the world, which is not always possible because the availability of maps and its number of areas. Different strategies are followed in order to face this problem, as an example a common practice is to use a map with a large number of areas and algorithmically construct the desired number of maps with the desired number of areas. The problem with these approaches is that different parts of the world could have different topological characteristics and hence different results and conclusions could be drawn if the base map is not properly selected. This situation becomes more critical when reviewing the litera-

ture is found that those topological characteristics are not taken into account and the selection of the large map is influenced by the location where the researcher applies; e.g; Duque et al. (2011a) uses maps from Colombia and United States to propose the regionalization algorithm MAX-P; Anselin and Moreno (2003) use United States maps to compare the performance of an statistical test over regular and irregular lattices; Mur Lacambra (1992) compares different methods to detect spatial autocorrelation over two artificial contiguity matrix and the map of Spain by regions. Fig. 3.1 shows the distribution of the Average number of neighbors, the first eigenvalue and the Aboav-Weaire coefficient of maps generated using United States by counties and Spain by NUTS3. In summary it is possible to conclude that maps of Spain are less connected and the number of neighbors are more correlated than maps of United States. As an example, selecting maps of United States instead of Spain could considerably increase the execution time of algorithms which its computational complexity increase with the number of neighbors (Aldstadt and Getis, 2006; Duque et al., 2010)

Because the difficulty of collecting an appropriate map and to take advantage of more controlled structures some authors use alternative approaches such as regular tilings or Voronoi diagrams. Regarding to regular lattices, they are easily generated, do not present restrictions about the number of areas and its over simplified structure usually make it easier to summarize mathematical theories and explain the behaviour of tested methodologies (Whittle, 1954; Bartlett, 1975; Griffith, 1987). The problem of using regular tilings is that they have some desirable theoretical properties that usually are not present in real maps and hence some assumptions or requirements need to be guaranteed before studding real phenomena using methodologies tested over them; e.g; the problem of spatial stationarity in STARMA models (Hooper and Hewings, 1981), improper conclusions of the properties of statistics in hypothesis testing, (Anselin and Moreno, 2003; Mur Lacambra, 1992), over qualification of computational efficiency of algorithms (Duque et al., 2010; Aldstadt and Getis, 2006). Regarding Voronoi diagrams, as will be shown in section 3.2, although they have random topological properties they are far from those of real maps.
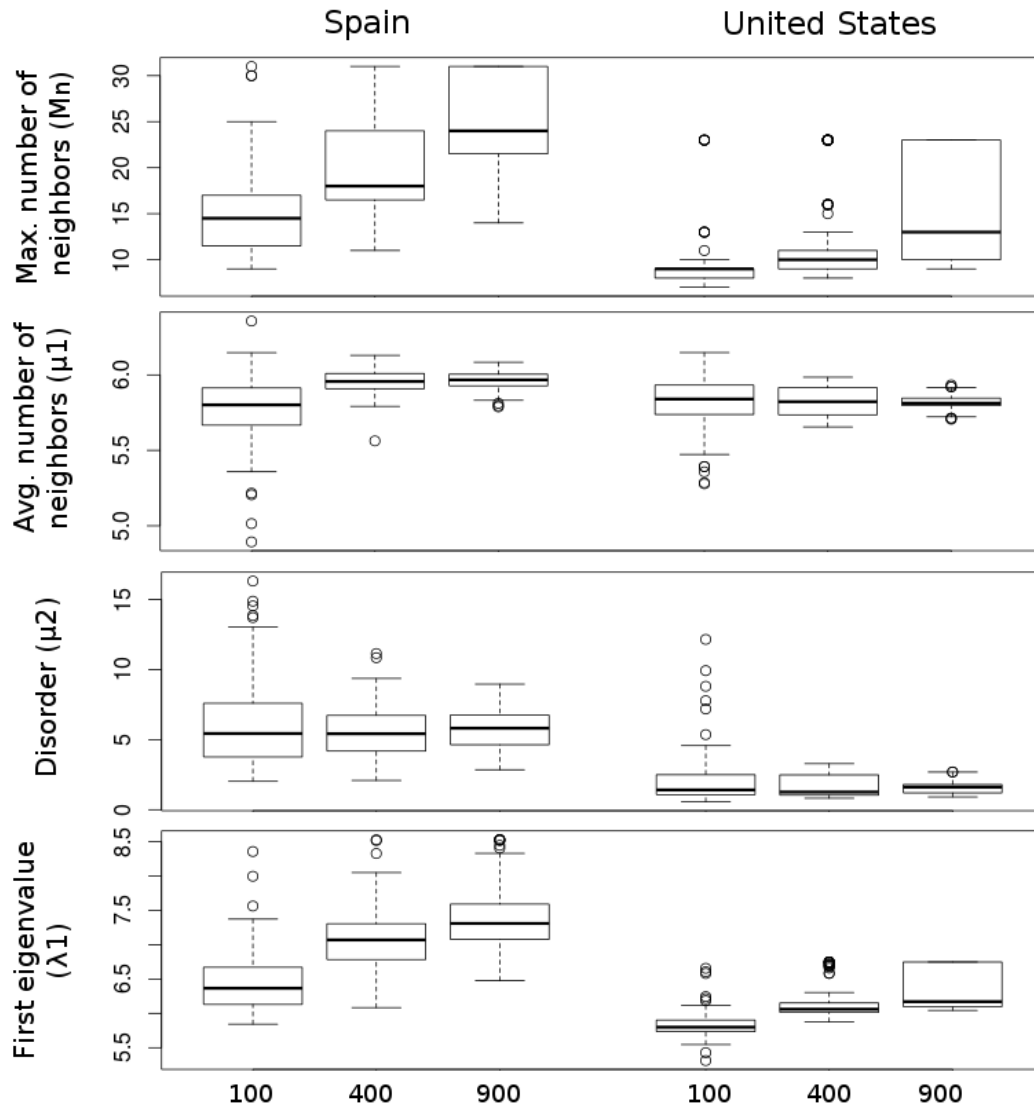
Fig. 3.1: Topological differences of maps from Spain and United States

3.2 Topological comparison of tilings

By now it is clear the advantages and disadvantages of the alternative approaches (regular tilings and Voronoi diagrams) to select a set of tilings to evaluate a methodology, however the question about if those categories of tilings are truly different remains open. To answer this, we propose a benchmark of topological indicators that a tiling must reproduce in order to be considered real and then compare the values of regular lattices and Voronoi diagrams. To obtain the bench-

mark the indicators proposed in section 2 are calculated over $1,000$ instances of maps with $100, 400, 900, 1,600, 2,500, 3,600$ areas, obtained from a sampling process over a world map at the smaller administrative division available in Hijmans et al. (2011), Fig. 3.2 illustrates the sampling process and Table 3.1 shows the number of maps sampled from each continental area. For Voronoi diagrams the same number of tilings with the same number of areas are used.[8] For regular lattices just one instance is enough. Table 3.2 shows the expected value of the indicators for each type of tiling grouped by number of areas, additionally, for real maps and Voronoi diagrams the standard deviation is calculated and presented bellow the expected values. In order to avoid the effect caused by the boundary to the values of M_n, m_n, μ_1, μ_2 the boundary areas are allowed to be neighbors of other areas but are not considered to calculate the indicators, however for S and λ_1 all areas are used because they are matrix properties.
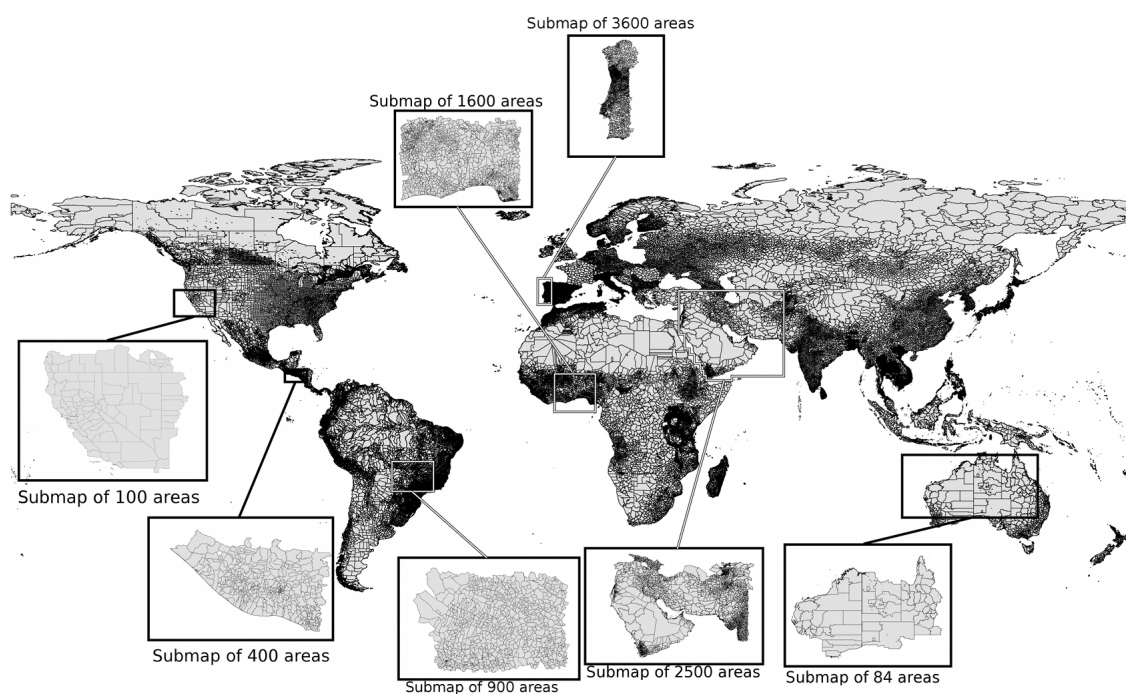
Fig. 3.2: Base map and example of a random irregular maps obtained from it.

For regular lattices it is evident that they are not capable to emulate the

---

[8]Voronoi diagrams where generated using random uniform points

Group	Number of areas	81	100	400	900	1,600	2,500	3,600
					Number of areas			
1: *America*	28,041	240	240	240	240	244	244	244
2: *Europa*	26,245	225	225	225	225	229	229	229
3: *Africa*	35,943	308	308	308	308	313	313	313
4: *Asia*	24,503	210	210	210	210	214	214	214
5: *Oceania*	1.880	17	17	17	17	0	0	0
Total	116,612	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Table 3.1: Number of maps from each group for a given number of areas.

topological characteristics of real maps in any of the indicators because they are absolutely ordered tilings, $\mu_2 = 0$, with 4 and 6 neighbors per area over the entire structure. Respect to λ_1 and S it is also important to note that regular tilings formed by hexagons are more connected than real maps while those formed by squares are less connected. Also note that in both cases the variations of the indicators while the number of areas increase is much lower than the variations of reals maps, which means that its topological characteristics are not highly affected by the size of the map.

About Voronoi diagrams it is found that they are not capable to generate atypically connected areas, values of M_n tending to 12 and m_n greater than 3. Regarding μ_1 they have a similar value of real maps, μ_1 close to 6 which is supported by the tenet of previous studies which suggest irregular tilings have an average number of neighbors close to 6 (Weaire and Rivier, 2009). From the values μ_2, close to 1.7, is possible to infer that they are more ordered than real maps. Additionally note that the magnitude of the change of the indicators while the size of the maps increase is smaller than real maps, showing that the topological characteristics of Voronoi diagrams are more stable for different sizes of maps. Finally the standard deviations evidence that the random properties of Voronoi diagrams are less disperse than on real maps.

		Number of areas						
		81	100	400	900	1,600	2,500	3,600
Real Maps	M_n	12.28 ±7.52	13.22 ±9.90	23.22 ±29.89	29.55 ±36.27	42.77 ±50.59	48.53 ±55.07	60.64 ±64.58
	m_n	2.33 ±1.11	2.13 ±1.06	1.55 ±0.86	1.23 ±0.60	1.04 ±0.22	1.01 ±0.10	1.00 ±0.00
	μ_1	5.57 ±0.65	5.59 ±0.61	5.67 ±0.49	5.69 ±0.45	5.70 ±0.46	5.72 ±0.37	5.72 ±0.37
	μ_2	5.85 ±13.85	6.72 ±22.58	9.76 ±28.79	7.90 ±15.35	8.85 ±12.82	7.73 ±9.39	8.00 ±8.11
	S	0.060 ±0.005	0.049 ±0.004	0.013 ±0.001	0.006 ±0.0005	0.003 ±0.0003	0.0022 ±0.00015	0.0015 ±0.0001
	λ_1	5.96 ±0.53	6.09 ±0.65	6.89 ±1.52	7.30 ±1.82	8.03 ±2.42	8.33 ±2.62	8.92 ±3.02
Reg. Lattice *(Squares)*	M_n	4	4	4	4	4	4	4
	m_n	4	4	4	4	4	4	4
	μ_1	4	4	4	4	4	4	4
	μ_2	0	0	0	0	0	0	0
	S	4.44	3.64	0.95	0.43	0.24	0.16	0.11
	λ_1	3.80	3.84	3.96	3.98	3.99	3.99	3.99
Reg. Lattice *(Hexagons)*	M_n	6	6	6	6	6	6	6
	m_n	6	6	6	6	6	6	6
	μ_1	6	6	6	6	6	6	6
	μ_2	0	0	0	0	0	0	0
	S	6.30	5.19	1.39	0.64	0.36	0.23	0.16
	λ_1	5.55	5.62	5.88	5.94	5.96	5.97	5.98
Voronoi Diagrams	M_n	9.15 ±0.77	9.36 ±0.79	10.37 ±0.75	10.90 ±0.74	11.26 ±0.70	11.49 ±0.67	11.71 ±0.68
	m_n	3.36 ±0.48	3.26 ±0.44	3.00 ±0.03	3.00 ±0.00	3.00 ±0.00	3.00 ±0.00	3.00 ±0.03
	μ_1	5.75 ±0.07	5.77 ±0.05	5.88 ±0.02	5.92 ±0.01	5.94 ±0.00	5.95 ±0.00	5.96 ±0.00
	μ_2	1.68 ±0.31	1.70 ±0.27	1.75 ±0.13	1.76 ±0.09	1.76 ±0.07	1.77 ±0.05	1.77 ±0.04
	S	6.67 ±0.08	5.47 ±0.05	1.44 ±0.00	0.65 ±0.00	0.37 ±0.00	0.24 ±0.00	0.17 ±0.00
	λ_1	5.88 ±0.05	5.96 ±0.05	6.20 ±0.03	6.26 ±0.02	6.28 ±0.02	6.29 ±0.02	6.30 ±0.02

Table 3.2: Average connectedness characteristics for real maps, regular lattices and Voronoi diagrams

4. Generating realistic tilings

In this section we propose a procedure to generate polymorphism irregular aperiodic asymmetric tilings with similar characteristics of real maps. To attain this we divide the section in four parts, the first and second parts briefly introduce the theoretical concepts which support the algorithms, the third part introduces a novel way to generate irregular polygons based on a mean reverting process in polar coordinates and the fourth part proposes a recursive procedure to generate tilings with similar topological characteristics of real maps using the irregular polygons of the third part.

4.1 Introduction to the mean reverting process

In order to propose the $MR - Polygon$ algorithm it is necessary to briefly introduce some formal background to state the mean reverting process. Next some basic concepts about Brownian motion and stochastic differential equations are introduced. The following definitions and concepts are constructed based on Mao (1997).

4.1.1 Brownian motion

Definition 4.1.1. *Let $(\omega, \mathfrak{F}, P)$ a probability space with a filtration $\{\mathfrak{F}\}_{t \geq 0}$. A standard unidimensional Brownian motion $\{B_t\}_{t \geq 0}$ is a continuous real valued process $\mathfrak{F}_t - adapted$ with the following properties.*

1. $B_0 = 0$

2. $\{B_t\}_{t \geq 0}$ *has stationary increments, which means that $B_t - B_s$ is identically distributed to $B_{t+h} - B_{s+h}$ for each $s, t \in T$ and each $s + h, t + h \in T$*

3. $\{B_t\}_{t \geq 0}$ *has independent increments, which means that for each $t_i \in T$ with $t_1 < t_2 < \cdots < t_n$ and $n \geq 1$, $B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent random variables.*

4. *For $0 \le s \le t < \infty$, the increment $B_t - B_s$ follows a normal distribution with mean 0 and variance $t - s$*

5. *$\{B_t\}_{t \ge 0}$ is continuous and not differentiable at any point*

A Brownian motion has other important properties that are important to mention. Among them one of the most important is the *self similarity* which means that a Brownian motion preserves, at different scales, the patterns of a Brownian instance. It is important to note that using this concept it is possible to establish an important relation between Brownian motions and fractals (see section 4.2).

1. A Brownian motion is a Gaussian process

2. The covariance of a Brownian motion is $cov(B_s, B_t) = min(s, t)$

3. A Brownian motion is $0.5 - selfsimilar$ which means

$$(T^{\frac{1}{2}} B_{t_1}, \ldots, T^{\frac{1}{2}} B_{t_n}) \stackrel{d}{=} (B_{T_{t_1}}, \ldots, B_{T_{t_n}})$$

4.1.2 Stochastic differential equations

In this section a brief introduction to stochastic differential equations is made. To attain this, first a general n-dimensional stochastic differential equation and its solutions are showed. Second the theorems of existence and uniqueness of solutions are stated. Third the Linear Stochastic Differential Equation is presented. Finally the solution of the linear equations in strict sense are presented and related with the mean reverting process used to propose the *MR-Polygon* algorithm.

Let $(\omega, \mathfrak{F}, p)$ be a complete probability space with a filtration $\mathfrak{F} \ge 0$. Let $B(t) = (B_1(t), B_2(t), \ldots, B_m(t))^T, t \ge 0$ an m-dimensional Brownian motion over the probability space. Let $t_0 < T$ between $0, \infty$. Let x_0 be an $\mathfrak{F}_{t_0}$-measurable random variable such that $E[x_0]^2 < \infty$. Let $f : \mathbb{R}^2 \times [t_0, T] \to \mathbb{R}^n$ and $g : \mathbb{R}^n \times [t_0, T] \to \mathbb{R}^{n \times m}$, both of them Borel-measurable.

Consider the n-dimensional Stochastic Differential Equation given by (4.1) with initial value $x(t_0) = x_0$. By definition this (4.1) is equivalent to the stochastic integral equation (4.2).

$$dx(t) = f(x(t), t)dt + g(x(t), t)dtB(t); t_0 \leq t < T \tag{4.1}$$

$$x(t) = x_0 + \int_{t_0}^t f(x(s), s)ds + \int_{t_0}^t g(x(s), s)dB(s); t_0 \leq t < T \tag{4.2}$$

Definition 4.1.2. *A R^n valuated stochastic process $\{x(t)\}_{t_0 \leq t \leq T}$ is known as solution of (4.1) if the following conditions are true:*

I) *$\{x(T)\}$ is continuous and $\mathfrak{F}_t$ adapted*

II) *$\{f(x(t), t)\} \in \mathscr{L}^1([t_0, T]; \mathbb{R}^n)$ and $\{g(x(t), t)\} \in \mathscr{L}^2([t_0, T]; \mathbb{R}^{n \times m})$*

III) *Equation (4.2) is true for each $t \in [t_0, T]$*

It is said that $\{x(t)\}$ is unique if any other solution $\{\bar{x}(t)\}$ is exactly the same as $\{x(t)\}$.

4.1.2.1 Existence and Uniqueness of solutions

Theorem 4.1.3. *Suppose the existence of two positive constants K and $\bar{K}$ that meets the next conditions, then exists a unique solution of equation (4.1) and it belongs to $\mathscr{M}^2([t_0, T]; \mathbb{R}^n)$.*

Lipschitz criterion *For every $x, y \in \mathbb{R}^n$ and $t \in [t_0, T]$*

$$|f(x, t) - f(y, t)|^2 \vee |g(x, t) - g(y, t)|^2 \leq \bar{K}|x - y|^2$$

Lineal growth condition *For every $(x, t) \in \mathbb{R}^n \times [t_0, T]$*

$$|f(x, t)|^2 \vee |g(x, t)|^2 \leq K(1 + |x|^2)$$

Theorem 4.1.4. *Suppose that the linear growth conditions are met, but the Lipschitz criterion is replaced by the **Local Lipschitz Criterion**:*

For each integer number $n \geq 1$, exists a positive constant K_n such that for every $t \in [t_0, T]$ and for every $x, y \in \mathbb{R}^n$ with $|x| \vee |y| \leq n$

$$|f(x,t) - f(y,t)|^2 \vee |g(x,t) - g(y,t)|^2 \leq K_n |x - y|^2$$

Then exists a unique solution $x(t)$ to (4.1) in $\mathscr{M}^2([t_0, T]; \mathbb{R}^n)$.

Theorem 4.1.5. *Suppose that the Lipschitz local criterion is met, but the lineal growth condition is replaced by the **Monotone condition**:*

Exists a positive constant K such that for every $(x, y) \in \mathbb{R}^n \times [t_0, T]$

$$x^t f(x,t) + \frac{1}{2}|g(x,t)|^2 \leq K(1 + |x|^2)$$

Then exists a unique solution $x(t)$ to (4.1) in $\mathscr{M}^2([t_0, T]; \mathbb{R}^n)$.

On one hand the Lipschitz local criterion ensures that the solution exists in $[t_0, \tau_\infty]$, where $\tau_\infty = \lim_{n \to \infty} \tau_n$. On the other hand the monotone condition ensures that $\tau_\infty = T$, which means that the solution exists in the interval $[t_0, T]$.

Consider the Stochastic Differential Equation given by (4.3) with initial value $x(T_0) = x_0$. If the conditions of existence and uniqueness are met for every finite subinterval $[t_0, T]$ into $[t_0, \infty]$, then it has a unique solution, called **global solution**, into $[t_0, \infty]$.

$$dx(t) = f(x(t), t)dt + g(x(t), t)B(t) \tag{4.3}$$

Theorem 4.1.6. *Suppose that for each real number $T > t_0$ and for each integer $n \geq 1$, exists a positive constant $K_{T,n}$ such that for every $t \in [t_0, T]$ and every $x, y \in \mathbb{R}^n$ with $|x| \vee |y| \leq n$ then $|f(x,t) - f(y,t)|^2 \vee |g(x,t) - g(y,t)|^2 \leq K_{T,n}|x - y|^2$. Also suppose that for every $T > t_0$, exists a positive constant K_T such that for every $(x, t) \in \mathbb{R}^n \times [t_0, T]$,*

$$x^T f(x,t) + \frac{1}{2}|g(x,t)|^2 \leq K_T(1 + |x|^2) \tag{4.4}$$

Then exists a unique global solution $x(t)$ of (4.3) and it belongs to $\mathscr{M}^2([t_0, T]; \mathbb{R}^n)$

4.1.2.2 Lineal Stochastic Differential Equations

Consider the n-dimensional stochastic differential equation given by (4.5) over $[t_0, T]$ where $F(.), G(.)$ are matrix valued functions of $n \times n$, $f(.), g(.)$ are functions with values in $\mathbb{R}^n$ and $B(t) = (B_1(t), B_2(t), \ldots, B_m(t))^T$ is a m-dimensional Brownian motion. It is said that (4.5) is homogeneous if $f(t) = g(t) = \cdots = 0$, it is said lineal in strict sense if $G_1(t) = G_2(t) = \cdots = 0$ and it is said autonomous if F, f, G_k, g_k are t independents. Also it is assumed that F, f, G_k, g_k are Borel-measurable and bounded over $[t_0, T]$. Under this framework, due to the existence and uniqueness theorem, equation (4.5) has unique solution in $\mathscr{M}^2([t_0, T]; \mathbb{R}^n)$ for each initial value $x(t_0) = x_0$.

$$dx(t) = (F(t)x(t) + f(t))dt + \sum_{k=1}^{m}(G_k(t)x(t) + g_k(t))dB_k(t) \qquad (4.5)$$

In this thesis we are only interested in linear equations in strict sense because, as be shown latter, the Mean Reverting process used to propose the *MR-Polygon* is included in this category. Consider the Linear Stochastic Differential equation in strict sense given by (4.6) over $[t_0, T]$ with initial value $x(t_0) = x_0$.

$$dx(t) = (F(t)x(t) + f(t))dt + \sum_{k=1}^{m}(g_k(t))dB_k(t) \qquad (4.6)$$

The linear homogeneous equation for (4.6) is the ordinary differential equation given by

$$dx(t) = F(t)x(t)dt, \qquad (4.7)$$

And the solution of (4.6) is given by

$$x(t) = \phi(t)\left(x_0 + \int_{t_0}^{t} \phi^{-1}(s)f(s)ds + \sum_{k=1}^{m} \int_{t_0}^{t} \phi^{-1}(s)g_k(s)dB_k(s)\right) \qquad (4.8)$$

Particularly when $F(t)$ is t independent, which means $F(t) = F$ is a constant matrix of $n \times n$ the fundamental matrix $\phi(t)$ has the form $\phi(t) = e^{F(t-t_0)}$ and it inverse $\phi^{-1}(t) = e^{-F(t-t_0)}$. Then equation 4.6 is solved by

$$x(t) = e^{F(t-t_0)} \left(x(t_0) + \int_{t_0}^{t} e^{-F(s-s_0)} f(s)ds + \sum_{k=1}^{m} \int_{t_0}^{t} e^{-F(s-s_0)} g_k(s)dB_k(s) \right) \quad (4.9)$$

A mean reverting process is a stochastic process that takes values following a long term tendency with presence of short term variations where the value of the process x at the moment t is given by the solution of the stochastic differential equation (4.10), where μ is the long term tendency, α is the mean reversion speed, σ is the gain on the diffusion term, $x(t_0)$ is the value of the process when $t = 0$ and $\{B_t\}_{t \geq 0}$ is an unidimensional Brownian motion. As can be seen (4.10) is formally a linear stochastic equation in strict sense over $[t_0, T]$, being $F(t) = -\alpha$, $f(x) = \alpha\mu$, $g_t = \sigma$ and is in just one dimension, which means $m = 1$.

$$dX_t = \alpha(\mu - X_t)d_t + \sigma dB_t \quad (4.10)$$

Under this framework, using (4.9), the solution of (4.10) with $\phi = e^{F(t-t_0)} = e^{-\alpha}$ is given by

$$x(t) = e^{-\alpha(t-t_0)} \left(x(t_0) + \int_{t_0}^{t} e^{\alpha(s-s_0)} \alpha\mu ds + \int_{t_0}^{t} e^{\alpha(s-s_0)} \sigma dB(s) \right), \quad (4.11)$$

Particularly to propose the *MR-Polygon* algorithm we can assume $t_0 = 0$ and then

$$
\begin{aligned}
x(t) &= e^{-\alpha t}\left(x_0 + \int_0^t e^{\alpha s}\alpha\mu ds + \int_0^t e^{\alpha s}\sigma dB(s)\right) \\
&= x_0 e^{-\alpha t} + e^{-\alpha t}\int_0^t e^{\alpha s}\alpha\mu ds + e^{-\alpha t}\int_0^t e^{\alpha s}\sigma dB(s) \\
&= x_0 e^{-\alpha t} + \frac{e^{-\alpha t}e^{\alpha t}\alpha\mu}{\alpha} + \int_0^t e^{-\alpha(t-s)}\sigma dB(s) \\
&= x_0 e^{-\alpha t} + \mu + \int_0^t e^{-\alpha(t-s)}\sigma dB(s).
\end{aligned} \tag{4.12}
$$

Evaluating (4.12) when $t_0 = 0$ is obtained the value of x_0

$$
x(0) = x_0 + \mu \rightarrow x_0 = x(0) - \mu,
$$

Substituting again into (4.12) the solution is given by

$$
x(t) = \mu + (x(0) - \mu)e^{-\alpha t} + \mu + \int_0^t e^{-\alpha(t-s)}\sigma dB(s).\forall t \in [0, T] \tag{4.13}
$$

For practical purposes generating polygons the use Euler approximation given by (4.14) is used, where $\epsilon_t \sim N(0,1)$,

$$
X_t = X_{t-1} + \alpha(\mu - X_{t-1})\Delta_t + \sigma\sqrt{\Delta_t}\epsilon_t \tag{4.14}
$$

4.2 Introduction to the fractal dimension

In this section the concept of fractal dimension as a way to measure a subjective feeling of how densely a fractal occupies a metric space is introduced, understanding a fractal as rough fragmented geometric shape which could be infinitely divided into scalable parts (Mandelbrot, 1982). The fractal dimension is a very important measure because it is a way to connect fractals with real world data, for example it is possible to get the fractal dimension of a coastline, an irregular polygon, a cloud, a tree or the colors emitted by the sun. Next the fractal dimension is

formally introduced, following Barnsley (2000).

Let $(\mathbb{X}, d)$ be a complete metric space. Let $A \in \mathscr{H}$ be a nonempty compact subset of $\mathbb{X}$, being $\mathscr{H}$ the space whose points are the compact subsets of $\mathbb{X}$, other than the empty set. Let $\epsilon > 0$. Let $B(x, \epsilon)$ denote the closed ball of radius ϵ centered at $x \in \mathbb{X}$. Define $\mathcal{N}(A, \epsilon)$ the least number of closed balls of radius ϵ needed to cover the set A. In other words,

$$\mathcal{N}(A, \epsilon) = \text{the smallest positive integer } M \text{ such that } A \subset \bigcup_{n=1}^{M} B(X_n, \epsilon),$$

For a set of points $x_n : n = 1, 2, 3, \ldots M \subset \mathbb{X}$. To prove that $\mathcal{N}(A, \epsilon)$ exists, surround each point $x \in A$ by an open ball of radius ϵ covering A using opens sets. Due to the compactness of A that cover has a finite sub-cover of $\hat{M}$ open balls. The closure of those balls we have a cover of $\hat{M}$ closed balls of radius ϵ. Let C the covers of A with at most $\hat{M}$ balls of radius ϵ, then C contains at least one element. Let $f : C \to 1, 2, 3, \ldots, \hat{M}$ be defined by $f(c) = $ number of balls in the cover $c \in C$. Then $f(c) : c \in C$ is a set of positive integers having at least an integer $\mathcal{N}(A, \epsilon)$.

It is said that a set A has fractal dimension D if for some positive number C $\mathcal{N}(A, \epsilon) \approx \epsilon^{-D}$. Where $\approx$ is used as follows. Let $f(\epsilon)$ and $g(\epsilon)$ be real valued functions, then it is said that $f(\epsilon) \approx g(\epsilon)$ if $\lim_{\epsilon \to 0} \frac{\log(f(\epsilon))}{\log(g(\epsilon))} = 1$. Solving for D then

$$
\begin{aligned}
D &\approx \frac{\log(\mathcal{N}(A, \epsilon)) - \log(C)}{\log(\frac{1}{\epsilon})} \\
&\approx \frac{\log(\mathcal{N}(A, \epsilon))}{\log(\frac{1}{\epsilon})} - \frac{\log(C)}{\log(\frac{1}{\epsilon})}
\end{aligned}
\tag{4.15}
$$

Using the definition of "$\approx$" it is clear that the second term in the right hand side of (4.15) tends to zero and then (4.16).

$$
D \approx \frac{\log(\mathcal{N}(A, \epsilon))}{\log(\frac{1}{\epsilon})}
\tag{4.16}
$$

Definition 4.2.1. *Let $A \in \mathcal{H}(\mathbb{X})$ where $(\mathbb{X}, d)$ is a metric space. For each $\epsilon > 0$ needed to cover A. If (4.17) exists, then it is called the fractal dimension of A.*

$$D = \lim_{\epsilon \to 0} \left(\frac{\log(\mathcal{N}(A, \epsilon))}{\log(\frac{1}{\epsilon})} \right) \tag{4.17}$$

In our case we are interested in the fractal dimension of the irregular boundary of a polygon, in this case then A is the boundary, ϵ is the length of the measurement instrument, and $\mathcal{N}(A, \epsilon)$ is the length of A using a measurement instrument of length ϵ, denoted as $L(A, \epsilon)$. Under this framework the fractal dimension is restated as (4.18).

$$D = \lim_{\epsilon \to 0} \left(\frac{\log((L(A, \epsilon))}{\log(\frac{1}{\epsilon})} \right) \tag{4.18}$$

4.3 Mean reverting areas (MR-Polygons)

4.3.1 Algorithm

The problem of characterizing the shape of irregular polygons, as those that comes from processes of urban growth or administrative borders, is commonly faced in two ways, evaluating its similitude with a circle (Haggett, 1977) or describing its boundary roughness with the fractal dimension (Batty and Longley, 1994; Frankhauser, 1998). Chen (2011) established a relationship between booth ways. In our case both of them are combined, the first approach to propose a way to create irregular polygons and the second to parametrize them under the concept of fractal dimension.

To compare irregular polygons with circles different classic indexes are used such as elongation ratio (Weeitty, 1969), form ratio (Horton, 1932), circularity ratio (Miller, 1953), compactness ratio (Gibbs, 1961; Cole, 1964; Richardson, 1961), ellipticity index (Stoddart, 1965) and the radial shape index (Boyce and Clark, 2012). According to Chen (2011) these indexes are based on the minimum circumcircle and the circle with equal area to the irregular polygon. Under this
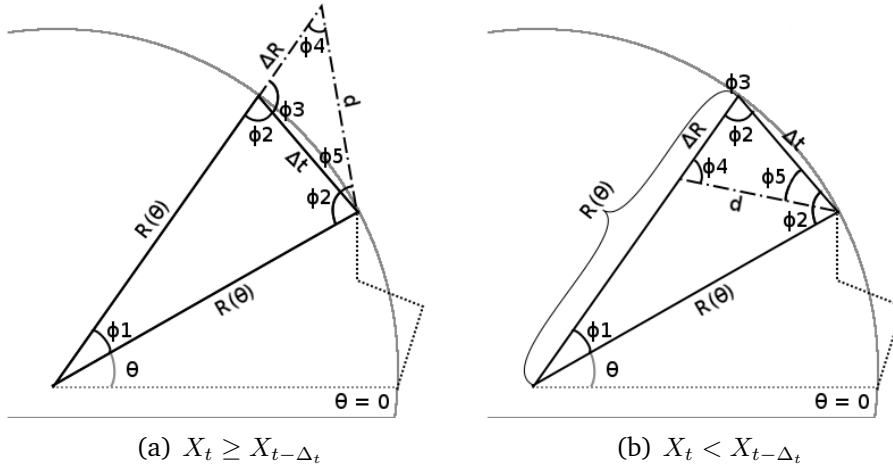
(a) $X_t \geq X_{t-\Delta_t}$ (b) $X_t < X_{t-\Delta_t}$

Fig. 4.1: Geometric problem to preserve the length and the fractal dimension of mean reverting process when is used to create an irregular polygon

approach an irregular polygon can be interpreted as an irregular boundary which follows a circle with a representative random noise. This interpretation guides us to use a mean reverting process, equation (4.14), to create irregular polygons.

Following the procedure showed in Pseudocode 1 it is possible to construct an irregular polygon P in polar coordinates where the points follows a circle with radius μ and where the distance, following the process, between any two points of X is exactly the same as the distance, following the frontier in counterclockwise direction, between their equivalent points in P. Angle ϕ_1 and Δ_r of Pseudocode 1 are the result of solving the geometric problem showed in Fig. 4.1 in order to define (4.19). Points of P are denoted as P_θ with θ between 0 and 2π. Regarding the effect of the parameters of X on P; α is the speed at which the process revert to the circle with radius μ, σ is the scaling factor of the irregularity of the polygon, X_0 is the initial radius. Δ_t can not be directly interpreted but it is affecting the magnitude of the angular steps of P.

$$P_{\theta+\phi_1} = \begin{cases} P_\theta + \Delta_R & \text{if } X_{t+\Delta_t} \geq Xt \\ P_\theta - \Delta_R & \text{if } X_{t+\Delta_t} < Xt. \end{cases} \tag{4.19}$$

Pseudocode 1: MEANREVERTINGAREA

$PARAMETERS:$

α : Mean revertion speed

σ : Noise scaling factor

μ : Long term tendence

X_0 : Initial point

Δ_t : Time step

$X_{t-\Delta_t} = X_0,$ Initial point of the mean reverting process

$P = [(0, X_0)],$ Irregular polygon in polar coordinates

while $\theta < 2\pi$

$\quad$ **do** $\begin{cases} \epsilon_t \leftarrow \text{N(0,1)} \\ X_t = X_{t-\Delta_t} + \alpha(\mu - X_{t-\Delta_t})\Delta_t + \sigma\sqrt{\Delta_t}\epsilon_t \\ d \leftarrow \text{distance between } X_t \text{ and } X_{t-\Delta_t} \\ R_\theta \leftarrow \text{Last radius of the irregular polygon} \\ \phi_1 = \arccos\left(\frac{2R_\theta - \Delta_t^2}{2R_\theta^2}\right) \\ \textbf{if } Xt \geq X_{t-\Delta_t} \\ \quad \textbf{then } \left\{\Delta_R = d\left(\cos\left(\arcsin\left(\frac{\Delta_t}{d}\cos\left(\frac{\phi_1}{2}\right)\right)\right) - \sin\left(\frac{\phi_1}{2}\right)\frac{\Delta_t}{d}\right)\right. \\ \quad \textbf{else } \left\{\Delta_R = -d\left(\cos\left(\arcsin\left(\frac{\Delta_t}{d}\cos\left(\frac{\phi_1}{2}\right)\right)\right) + \sin\left(\frac{\phi_1}{2}\right)\frac{\Delta_t}{d}\right)\right. \\ R_{\theta+\phi_1} = R_\theta + \Delta_R \\ \textbf{Add } (\theta + \phi_1, R_{\theta+\phi_1}) \textbf{ to } P \\ \textbf{increase } \theta \textbf{ in } \phi_1 \end{cases}$

Replace last point of P **to** $(0, X_0)$

return (P)

4.3.2 Parametrization of MR-Polygons

The second approach to characterize irregular polygon is through the fractal dimension D of its irregular boundary, which is a number between 1, for smooth boundaries, and 2 for rough boundaries. In general, an object is considered as fractal if it is dotted with irregular characteristics that are present at different scales of study (Mandelbrot, 1982). The concept of fractal dimension of boundaries refers to Richardson (1961) who infers that the length L of an irregular boundary, like a coast, is given by (4.20) where ϵ is the length of the measurement instrument, $\hat{C}$ is a constant and D is the fractal dimension of the irregular shape. For practical purposes D is obtained using (4.21) and is given by 1 minus the slope of $\log(L(\epsilon))$, this procedure is commonly known as the Richardson plot. For the formal framework of the fractal dimension see section 4.2. Note that (4.21) uses $\hat{C}$ and $(1 - D)$ instead of the constant C and $-D$ as the formal definition, this is obtained making $\hat{C} = \frac{C}{\epsilon}$.

$$L(\epsilon) \;=\; \hat{C}\epsilon^{1-D} \tag{4.20}$$

$$\log(L(\epsilon)) \;=\; (1-D)\log(\epsilon) - \log(\hat{C}) \tag{4.21}$$

Practically in all cases the Richardson plot can be explained with two line segments having different slopes, then two fractal dimensions can be obtained; textural, for small scales, and structural, for large scales (Kindratenko and Treiger, 1996). As an example, Fig. 4.2, shows a segment of the United States east coast taken from Google maps on two resolutions. Note that while the resolution increase some irregularities that were imperceptible on low resolution become visible. In this sense it may be said that irregularities at low resolution define the general shape and are related with structural dimension, while the irregularities at high resolution give the noise and are related with textural dimension. Because of this is that regional scientists are used to use highly sampled maps which preserves the general shape but removes the little variations looking for some benefits without changing the topological configuration of maps (Douglas, 1973).
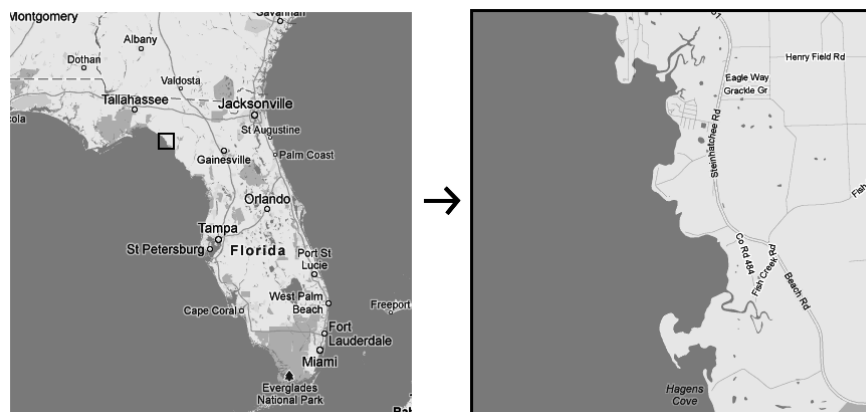

Fig. 4.2: Illustrative example of Irregularities explained by the Structural and Textural dimension.

In the field of stochastic signals some approaches based on different approximations of the length have been made in order to characterize them through its fractal dimension. In our case an experimental approach based on the fractal di-

mension of real polygons is proposed to select an appropriated combination of the parameters α and σ that generates realistic polygons, because our interest is on general shape more than little variations only structural dimension is taken into account.[9] The experiment is divided in two parts, in the first part the frequency histogram of fractal dimensions of the real polygons is constructed. In the second part we propose a range of possible values of α and σ, given μ, X_0, Δ_t, that generates fractal dimensions close to the obtained in the first part. Because the level of the long term tendency, μ, does not affect the length of X and Pseudocode 1 guarantee that length is preserved then μ can be defined as constant without affecting the fractal dimension, hereinafter it is assumed $\mu = X_0 = 10$. Regarding Δ_t it must be a small number in order to properly infer both fractal dimensions, from now on it is defined as 0.001.

The empirical distribution of the fractal dimension of the irregular polygons is calculated over a random sample of 10.000 polygons from the world map used in section 3. The result of this empirical distribution is showed in Fig. 4.3.
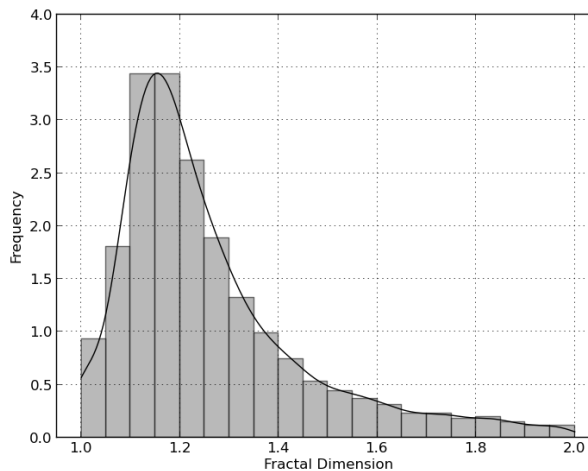
Fig. 4.3: Empirical frequency distribution of structural fractal dimensions of real polygons.

To summarize the fractal dimension of the stochastic polygons obtained with Pseudocode 1 a surface, of the average dimensions, is generated while both un-

---

[9]To calculate the structural dimension we use the EXACT procedure (Allen et al., 1995) with a is a small value of Δ_t, later both dimension were divided using a k-means clustering algorithm over the points of the Richardson plot.

known parameters, α and σ, changes from 0.01 to 5 whit steps of 0.1, Fig 4.4. Based on the surface it can be inferred that fractal dimension is mainly affected by changes to σ specially when looking at small dimensions, additionally it is possible to guarantee that fractal dimensions close to 1.23 are obtained when σ takes values between 1.2 and 1.5 regardless of the value of α. Fig 4.5 shows some examples of polygons using different values of α and σ. Note that polygons of the middle row, realistic combinations of α and σ, look similar between them which confirms the conclusions made about the influence of the parameters to the structural dimension, additionally in the same figure, both, original (gray line) and sampled (black line) polygons are showed reinforcing the fact that sampling a polygon does not affects the structural dimension. From now on and looking for computational efficiency only sampled polygons are used.
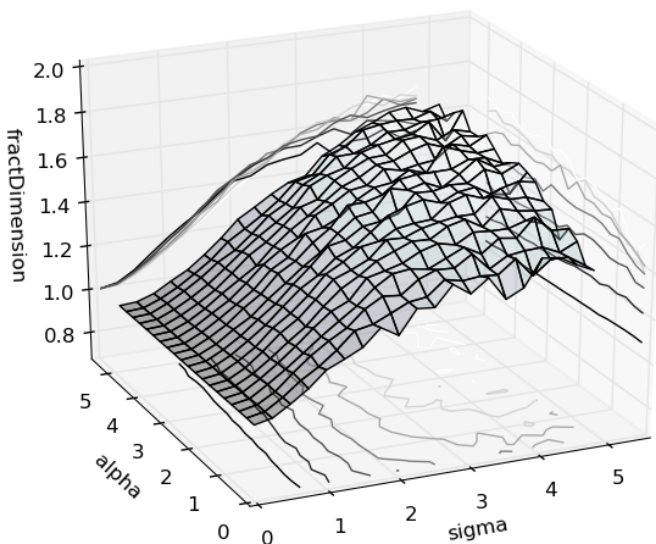
Fig. 4.4: Fractal dimension of simulated polygons in function of α and σ.

4.3.3   MR-Polygons with python

Using clusterpy is possible to create *MR-Polygons* with just a few lines of code. Next some examples are presented.[10] Due to the objective of this project no

---

[10]To install clusterpy please follow the instructions of section 8.1
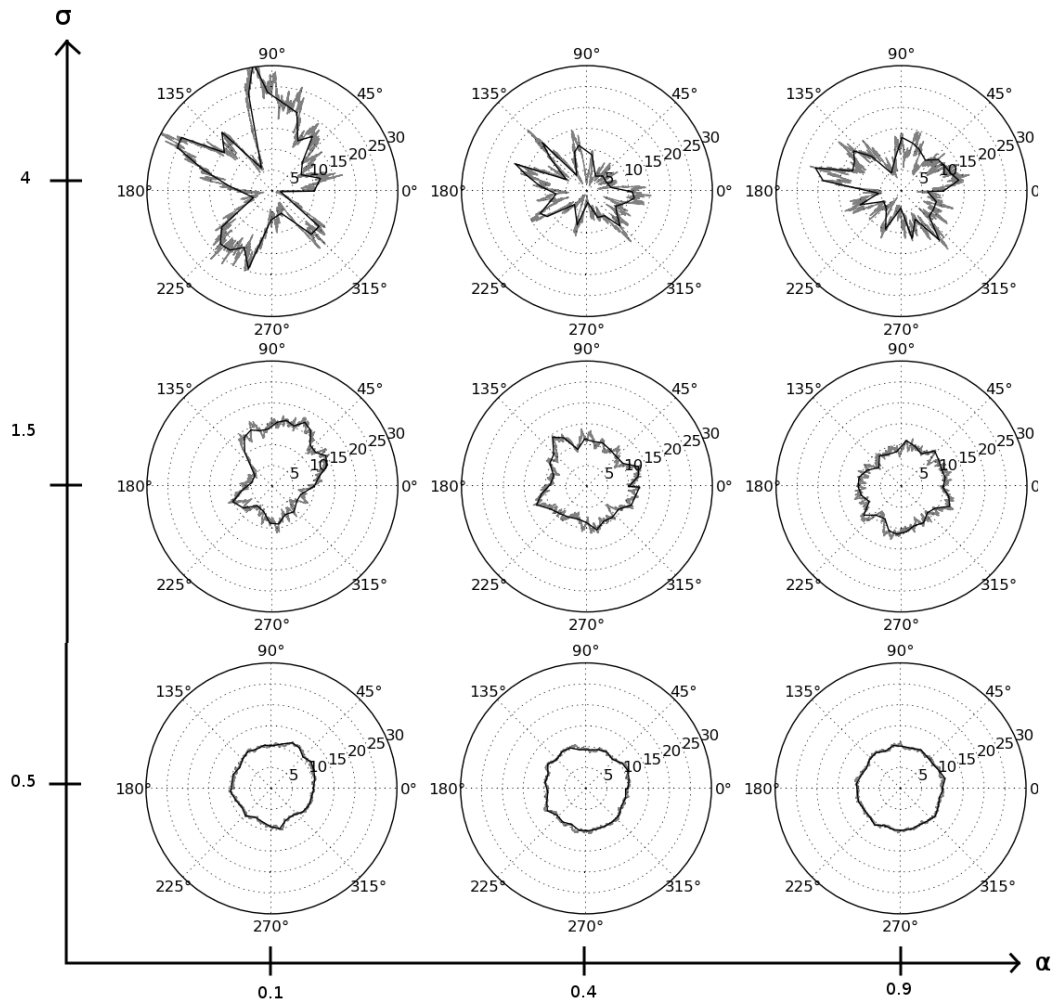
Fig. 4.5: **Examples of stochastic polygons generated using (Pseudocode 1) with different values of σ and α.**

visualization tool is exposed, however for more information see python site.

Basic *MR-Polygon* Creating a basic *MR-Polygon* using just α, σ and keeping the others parameters as default.

```
>>> import clusterpy
>>> from clusterpy.core.toolboxes import mrpolygon
>>> pol = mrpolygon(0.5, 1.2)
```

Complete *MR-Polygon* Creating a *MR-Polygon* definig $\alpha, \sigma, \mu, X_0, \Delta_t, nPoints$.

```
>>> import clusterpy
>>> from clusterpy.core.toolboxes import mrpolygon
>>> pol = mrpolygon(0.5, 1.2, mu=10, X_0=10, dt=0.001, nPoints=30)
```

Getting the coordinates of *MR-Polygon* The function `mrpolygon` return 6 elements into a python list [`non-sampled angles`, `non-sampled ratios`, `sampled angles`, `sampled ratios`, t, X_t]. To get the sampled polygon it is necessary to get the second and third elements of the result and zip them into a single list. At the end of the example the polar polygon is transformed to Cartesian coordinates

```
>>> import clusterpy
>>> from clusterpy.core.toolboxes.rimaps import mrpolygon, polarPolygon2cartesian
>>> pol = mrpolygon(0.5, 1.2)
>>> sampledAngles = pol[2]
>>> sampledRatios = pol[3]
>>> polarPolygon = zip(sampledAngles, sampledRatios)
>>> polarPolygon
[(0, 10), (0.23, 9.47), (0.44, 9.51), . . . , (6.28, 10)]
>>> cartePolygon = polarPolygon2cartesian(polarPolygon)
>>> cartePolygon
[(10, 0), (9.22, 2.18), (8.58, 4.10), . . . , (10, 0)]
```

4.4 Recursive Irregular maps (RI-Maps)

4.4.1 Algorithm

On section 3 it was showed that most commonly used tilings are not capable to emulate the topological characteristics of real maps highlighting the necessity of new tools to create realistic tilings. Looking to satisfy that necessity and based on the fractal characteristics of geographic objects, such as cities, exposed in Batty and Longley (1994) we conceptualize a recursive algorithm on which an irregular frontier is recursively divided using *MR-Polygons* which at the end of the recursion are returned to divide the original frontier into the desired number of areas. Our

conceptualization of the algorithm was made under three principles. i) *Scalability:* Preserving the complexity of the algorithm when the number of areas increases to make possible the construction of tilings with a large number of areas. ii) *Fractality:* Constructing maps from a hierarchically perspective on which each polygon is divided recursively, looking for the fractality observed in real maps which are divided on different scales i.e: United States is divided in States which in turn are divided in counties which also are divided in townships. iii) *Correlativity:* Encouraging the presence of spatial agglomerations of areas with similar sizes.

In Pseudocode 2 we present the *RI-Maps* algorithmic approach to create a polymorphism irregular aperiodic asymmetric with similar topological character-istics of real maps dividing an irregular polygon pol into n irregular polygons. As notation p_a is referring to the area of the polygon p. Because the recursive division of pol is made using *MR-polygons* then it is necessary to define the values of α and σ to be used, in this case and based on the experiments made in Section 4.3 they are taken from a uniform distribution between 0.1 and 0.5 for α, and 1.2 and 1.5 for σ. About μ X_0 and Δ_t they are defined with the same values established in Section 4.3. Regarding the sampling process to preserve the computational treatability of geocomputational operations *MR-Polygons* are sampled to 30 points.

$$\alpha \in [0.1, 0.5]$$

$$\sigma \in [1.2, 1.5]$$

As can be observed in the *RI-Maps* procedure there are three parameters that have not been explained yet (p_1, p_2, p_3). Regarding p_1 it is related with the scaling factor to be used over the *MR-Polygons* which divides the irregular frontiers, the higher value of p_1 the bigger irregular polygons to be used. About p_2 and p_3 they allow the presence of the spatial agglomerations of areas with similar sizes. p_2 define how probable it is to increase the number of polygons to create into a divisor polygon and p_3 is the percentage of the missing areas on which the divisor polygon is increased. Note that if p_2 or p_3 take values of zero then the map will be homogeneously divided into areas with similar size and no disorder will be

present.

For a better illustration of the parameter selection effect on the topological characteristics of *RI-Maps* some examples are presented in Table 4.1. These examples are presented in small pictures for easy comparison of the general topological differences, however in the following sections more detailed examples are presented. In the first row of the table p_2 and p_3 are defined as 0 and as can be observed generated tilings are very organized and do not show spatial correlation. The second and third combinations show spatial agglomerations of areas with similar sizes and are less ordered than those of the first row, however in the second row the groups of areas with similar sizes are less frequent and marked. As will be shown in section 4.4.2 the tilings generated with the third combination are the more realistic in terms of their topological characteristics.
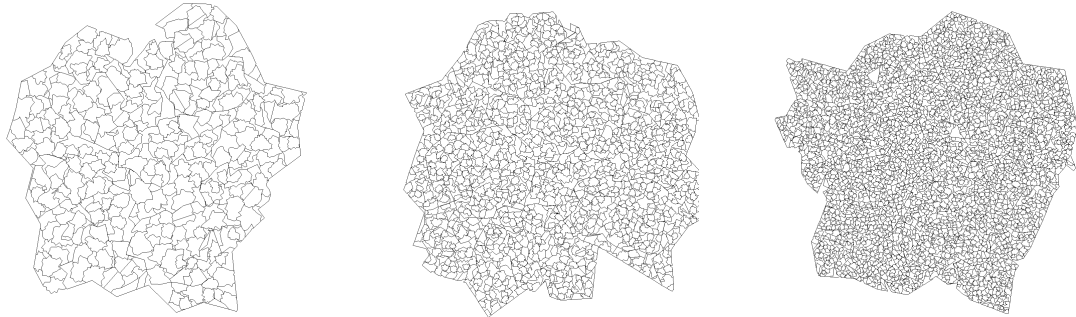
i) $p_1 = 0.1$, $p_2 = 0.0$, $p_3 = 0.0$

ii) $p_1 = 0.1$, $p_2 = 0.1$, $p_3 = 0.5$

iii) $p_1 = 0.02174$, $p_2 = 0.0676$, $p_3 = 0.3077$

Table 4.1: Examples of *RI-Maps* of 400, 1.600 and 3.600 areas using different combinations of parameters

---

Pseudocode 2: RECURSIVEIRREGULARMAP

---

if $n == 1$
  then $areas \leftarrow [pol]$
  else if $n == 2$
  then $\begin{cases} pol_1, pol_2 \leftarrow \text{Divide pol in 2} \\ areas \leftarrow [pol_1, pol_2] \end{cases}$

else $\begin{cases} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \mathbf{do} \begin{cases} \\ \end{cases} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{cases}$

$missingAreas \leftarrow n$
$uncoveredArea \leftarrow pol$
$coveredArea \leftarrow \text{Empty pol}$
$areas \leftarrow \text{Emtpy list}$
$scalingFactor \leftarrow \sqrt{\frac{p_1 \times pol_a}{n \times \pi \times \mu^2}}$
while $\frac{uncoveredArea_a}{pol_a} >= 0.03$

$uncovered2select \leftarrow \text{Bigger part of uncovered area}$
if $missingAreas \times \frac{uncovered2select_a}{uncoveredArea_a} <= 1.5$
  then $\begin{cases} \text{Append } uncovered2selet \text{ to } areas \\ coveredArea \leftarrow coveredArea \cup uncovered2slect \\ missingAreas \leftarrow missingAreas - n_i \end{cases}$
else $\begin{cases} \alpha_r \leftarrow \text{Random number between } \alpha_{min} \text{ and } \alpha_{max} \\ \sigma_r \leftarrow \text{Random number between } \sigma_{min} \text{ and } \sigma_{max} \\ poli \leftarrow \text{MEANREVERTINGAREA}(\alpha_r, \sigma_r, \mu, X_0, \Delta_t) \\ poli \leftarrow \text{Multiply each ratio of } pol_r \text{ by } scalingFactor \\ poli \leftarrow \text{Center } pol_r \text{ randomly into } uncovered2select \\ poli \leftarrow (pol_r - coveredArea) \cap pol \\ poli \leftarrow \text{Bigger part of } pol_r \\ n_i \leftarrow missingAreas \times \frac{poli_a}{uncoveredArea_a} \\ \textbf{if } Uniform(0,1) < p_2 \\ \quad \textbf{then } n_i = n_i + missingAreas \times p_3 \\ n_i \leftarrow \text{Round } n_1 \\ \textbf{if } n_i \geq 1 \\ \quad \textbf{then } \begin{cases} areasi \leftarrow \text{RECURSIVEIRREGULARMAP}(n_i, poli) \\ \text{Extend } areas \text{ with } areasi \\ coveredArea \leftarrow coveredArea \cup areasi \\ missingAreas \leftarrow missingAreas - n_i \end{cases} \end{cases}$
$uncoveredArea \leftarrow pol - coveredArea$

Append interior holes of $coveredArea$ to $areas$
$coveredArea \leftarrow \text{Union of } areas$
while Length of $areas < n$
  do $\{$**Append the smaller area to its bigger neighbor**
while Length of $areas > n$
  do $\{$**Divide the bigger area**

return $(areas)$

4.4.2 Parametrizing RI-Maps

To properly define p_1, p_2, p_3 to get realistic topological characteristics using *RI-Maps* we use a standard genetic algorithm, where the population γ of the iteration i, denoted as γ^i, is formed by the genomes $\gamma^i_j = [p^i_{j_1}, p^i_{j_2}, p^i_{j_3}]$, where $p^i_{j_1}, p^i_{j_2}, p^i_{j_3}$

are real numbers between 0 and 1 representing instances of p_1, p_2, p_3 denoted as phenomes. In this case $i \in \mathbb{N}$ between 0 and 20 and $j \in \mathbb{N}$ between 0 and 200. In order to evaluate the quality of each genome the fitness function is defined as (4.22), where θ is a set of number of areas, ϕ_k is the relative importance for a maps of k areas, and $f_k(\gamma_j^i)$ is a function given by (4.23) that measures how different are, in average, the topological indicators of *RI-Maps* formed by k areas using the phonome γ_j^i. Looking for simplicity stating (4.23) we denote the vector of real indicators as $\Psi_k = [M_n, m_n, \mu_1, \mu_2, S, \lambda_1]$ and $\Psi_k(\gamma_j^i)$ as the same vector for the mean values *RI-Maps* with k areas using γ_j^i, to refer the lth indicator Ψ_k^l and $\Psi_k^l(\gamma_j^i)$ is used. Regarding ns it is the number of simulations to be generated with each genome.

$$F(\gamma_j^i) = \frac{\left(\sum_{k \in \theta} \phi_k f_k(\gamma_j^i)\right)}{\sum_{k \in \theta} \phi_k} \tag{4.22}$$

$$f_k(\gamma_j^i) = \frac{\sum_{l=1}^{6} \frac{\left(\sum_{s=i}^{ns} \Psi_k^l(\gamma_j^i)\right) - ns\Psi_k^l}{ns\Psi_k^l}}{6} \tag{4.23}$$

To initialize the algorithm an initial random population of 100 genomes is generated and the four genomes with lower fitness function are obtained. The subsequent populations are formed by two parts, the first 64 genomes are all the possible combinations of the last best 4 genomes, the others 36 genomes are random modifications of those 64 genomes. Due to the computational time required to evaluate (4.22) only maps of 400 and 1.600 were used with an importance of $\phi_{400} = 1$ and $\phi_{1.600} = 2$ respectively. Fig. 4.6 shows the best fitness functions of each iteration. As conclusion the best phoneme is $p_1 = 0.02174$, $p_2 = 0.0676$, $p_3 = 0.3077$. As is observed the fitness function shows a quick decrease in the first 5 iterations to stay constant until $13th$ iteration and achieve the optimal value in the $15th$ iteration. In section 5, the topological indicators of *IR-Maps* with the optimal parameters are calculated as a complement of 3.2 and some examples are showed.
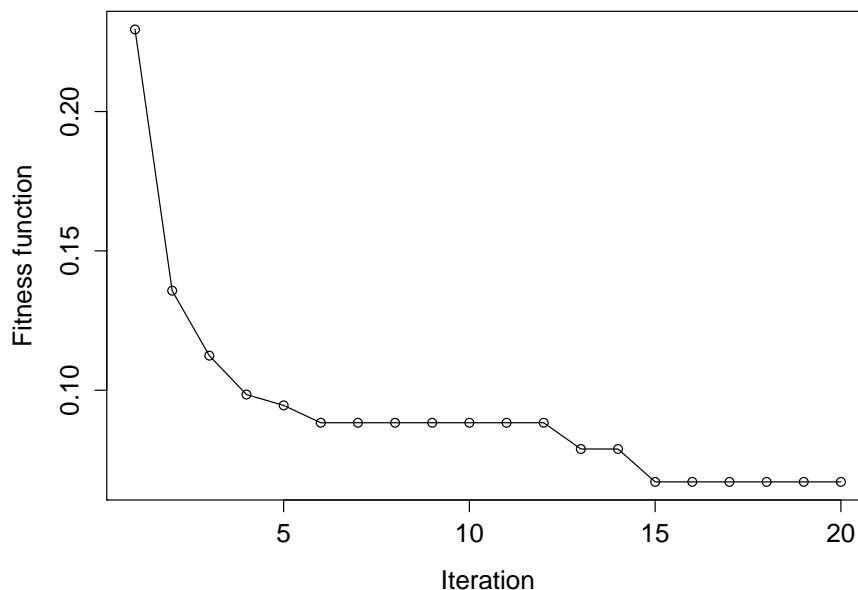
Fig. 4.6: Fitness function across the iterations

4.4.3 RI-Maps with python

Using clusterpy is possible to create *RI-Maps* and calculate the topological indicators of tilings such as real maps or simulated instances. Next some examples are presented.[11]

Basic *RI-Map* Creating and exporting to a shape file a basic *RI-Map* defining just the number of areas and keeping the others parameters as default.

```
>>> import clusterpy
>>> rimap = clusterpy.rimap(1600)
>>> rimap.exportArcData("<shapeFile>")
```

Complete *RI-Map* Creating and exporting to a shape file a *RI-Map* defining the following parameters, N_p, $[\alpha_{min}, \alpha_{max}]$, $[\sigma_{min}, \sigma_{max}]$, Δ_t, $p1$, $p2$, $p3$, $boundary$, where N_p is the number of points to sample of each polygon, $[\alpha_{min}, \alpha_{max}]$ and $[\sigma_{min}, \sigma_{max}]$ are the minimum and maximum values of α and σ to be used for the *MR-Polygons*, and $boundary$ is a path pointing to a map stored in a

---

[11]As in the previous sections it is assumed that python and clusterpy is installed.

shapefile with a single polygon to be used as initial irregular polygon. If any of the parameters is omitted then the value proposed in section 4.4 is used, if *boundary* is omitted then the initial polygon will be a *MR-Polygon*.

Example 1: Using a *MR-Polygon* as initial boundary

```
>>> import clusterpy
>>> rimap = clusterpy.rimap(1600,N=30,
      alpha=[0.1,0.5],sigma=[1.2,1.5],dt=0.1,pg=0.02174,pu=0.0675,su=0.307701)
>>> rimap.exportArcData("<shapeFile>")
```

Example 1: Using an existing irregular map as initial boundary

```
>>> import clusterpy
>>> rimap = clusterpy.rimap(1600,N=30,
      alpha=[0.1,0.5],sigma=[1.2,1.5],dt=0.1,pg=0.02174,pu=0.0675,su=0.307701),
      boundary="<boundary path>")
>>> rimap.exportArcData("<shapeFile>")
```

Topological indicators In order to obtain the topological indicators of a tiling using clusterpy is necessary to use the function `topoStats` which returns a list with 6 indicators in the following order $[M_n, m_n, \mu_1, \mu_2, S, \lambda_1]$.

Example 1: Getting the maximum number of neighbors and the sparseness for a regular lattice

```
>>> import clusterpy
>>> grid = clusterpy.createGrid(10,10)
>>> stats = grid.topoStats(regular=True)
>>> stats[0]
4
>>> stats[5]
0.03636
```

Example 2: Getting the maximum number of neighbors and the sparseness for a map stored in shapefile

```
>>> import clusterpy
>>> layer = clusterpy.importArcData("<shapefile>")
>>> stats = layer.topoStats()
>>> stats[0]
>>> stats[5]
```

Example 3: Getting the maximum number of neighbors and the sparseness for a *RI-Map*

```
>>> import clusterpy
>>> layer = clusterpy.rimap(1600)
>>> stats = layer.topoStats()
>>> stats[0]
23
>>> stats[5]
0.0035
```

5. Results

As was exposed in section 3.2 using Table 3.2 the topological characteristics of most common used tilings in regional sciences can not be considered realistic environments. In this section we present the topological characteristics of *RI-Maps* in the same tabular structure and using plots of the average values of irregular tilings, Voronoi diagrams, real maps and *RI-Maps*. At the end of the section a set of examples of *RI-Maps* are presented and based on the computational results of (Duque et al., 2010) a computational application is exposed.

5.1 Topological characteristics of *RI-Maps*

In Table 5.1 the average and standard deviation of the 6 indicators of *RI-Maps* using the optimal parameters ($p_1 = 0.02174$, $p_2 = 0.0675$, $p_3 = 0.3077$) are presented. Additionally in Fig. 5.1 the average values are plotted for easy comparison. To construct the indicators of *RI-Maps* a total of 100 instances were used for each size of map. In Fig. 5.1 it is observed that the maximum number of neighbors of *RI-Maps* is close to the value of real maps. Regarding the average number of neighbors, *RI-Maps* value is closer to Voronoi diagrams than real maps, however for practical purposes having values of μ_1 close to 6 or 5.2 is clearly not a big difference. Additionally it is important to note from μ_2 that *RI-Maps* are disordered than Voronoi diagrams making them more appropriate to emulate real maps. From Fig. 5.1 it is important to highlight the value of λ_1 close to those of real maps, specially for maps with more than 400 areas. Finally from the values of S clearly all the compared tilings show similar values. Finally from Table 5.1 it is found that standard deviations are small, which means that *RI-Maps* topological characteristics are less disperse than those of real maps.

5.2 *RI-Maps* examples

In order to complement the topological characteristics showed before three groups of *RI-Maps* examples are presented. In the first group Fig. 5.2 to 5.5 maps

41

	Number of areas						
	81	100	400	900	1,600	2,500	3,600
M_n	17.24 ±9.40	16.82 ±6.40	25.26 ±8.20	30.50 ±6.87	37.71 ±6.12	46.72 ±6.24	51.60 ±6.86
m_n	1.98 ±1.03	1.83 ±0.99	1.43 ±0.79	1.06 ±0.28	1.02 ±0.14	1.01 ±0.10	1.00 ±0.00
μ_1	5.52 ±0.25	5.59 ±0.22	5.82 ±0.07	5.87 ±0.05	5.90 ±0.04	5.91 ±0.04	5.92 ±0.04
μ_2	7.38 ±8.08	6.31 ±4.15	7.82 ±2.26	8.75 ±1.19	10.50 ±1.35	12.59 ±1.81	12.25 ±1.38
S	0.064 ±0.002	0.052 ±0.001	0.014 ±0.000	0.006 ±0.0000	0.004 ±0.0000	0.0022 ±0.00002	0.0016 ±0.00001
λ_1	6.64 ±0.69	6.70 ±0.47	7.50 ±0.55	7.87 ±0.44	8.39 ±0.37	9.00 ±0.38	9.28 ±0.44

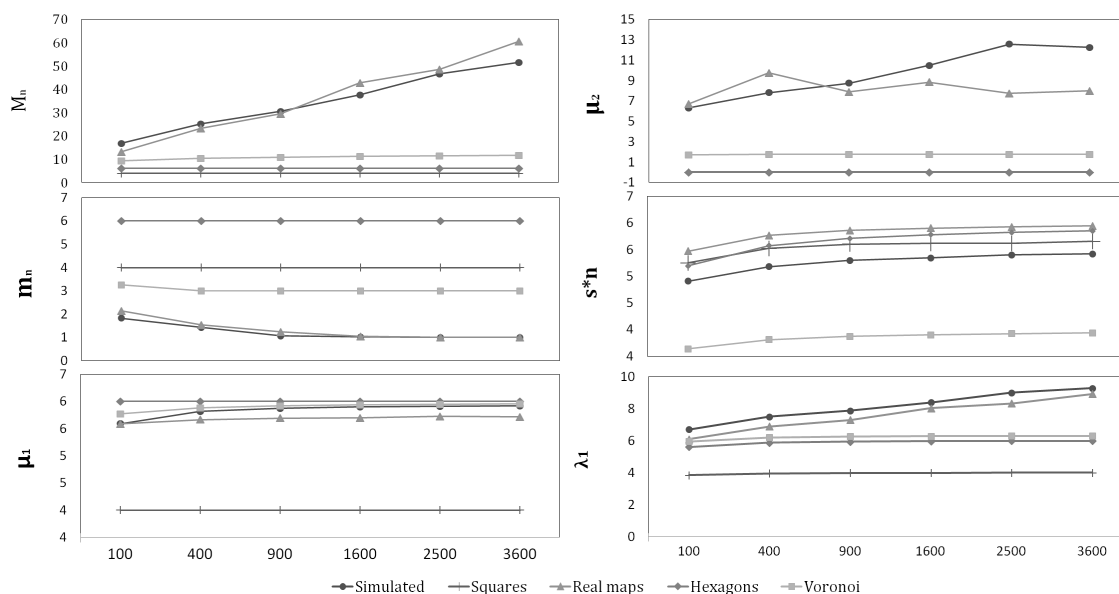**Table 5.1: Average connectedness characteristics for simulated maps**

Fig. 5.1: Comparison of topological characteristics of reals maps, *RI-Maps*, Voronoi diagrams and regular tilings.

with 400, 1.600, 3600, 7.200 are presented with an Spatial Autoregressive Process with $\rho = 0.7$.

In the second group of examples, Fig. 5.6 to 5.7, a set of maps, created taking advantage of the fact that *RI-Maps* can use any irregular polygon as the external boundary. In Fig 5.6(a) United States by counties is presented and Fig 5.6(b) shows the result of use main single part of the external boundary of United
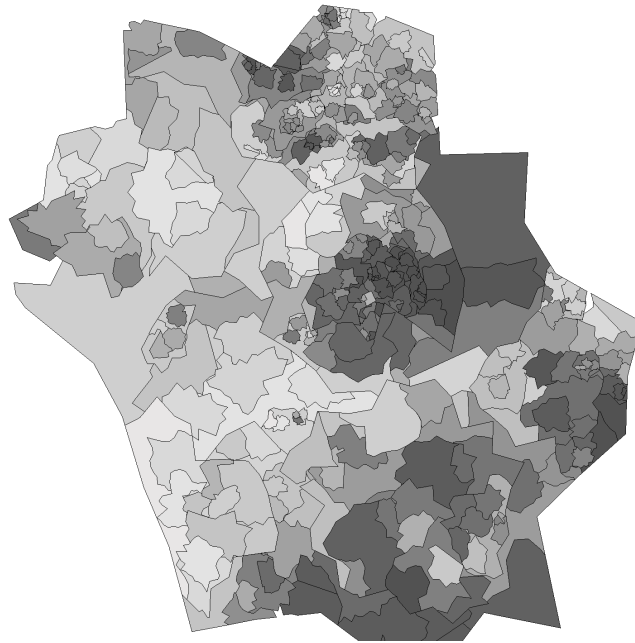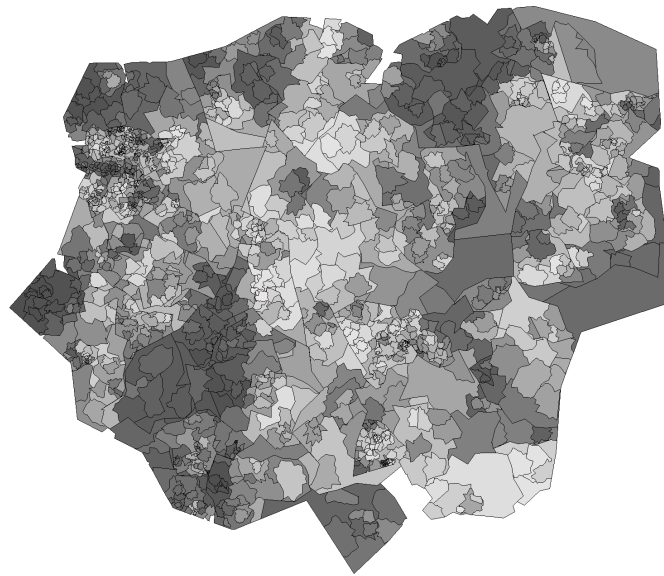
Fig. 5.2: SAR(0.7) process over a *RI-Map* of 400 areas

Fig. 5.3: SAR(0.7) process over a *RI-Map* of 1.600 areas
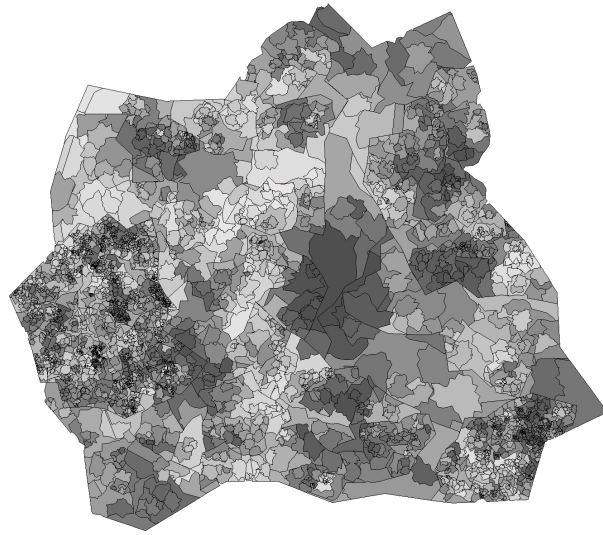
**Fig. 5.4:** SAR(0.7) process over a *RI-Map* of 3.600 areas

Fig. 5.5: SAR(0.7) process over a *RI-Map* of 7.200 areas

States as input to construct an *RI-Maps* with the same number of areas as counties into the single part (3304 polygons). In 5.7 and Fig 5.8 the same exercise is made using Colombia dividing by municipalities (1124 polygons) and Italy in third administrative level (7308 polygons). Finally to present the capacity of our algorithm of generating maps with a large number of areas a *RI-Map* of $20,000$ areas is presented in 5.9.
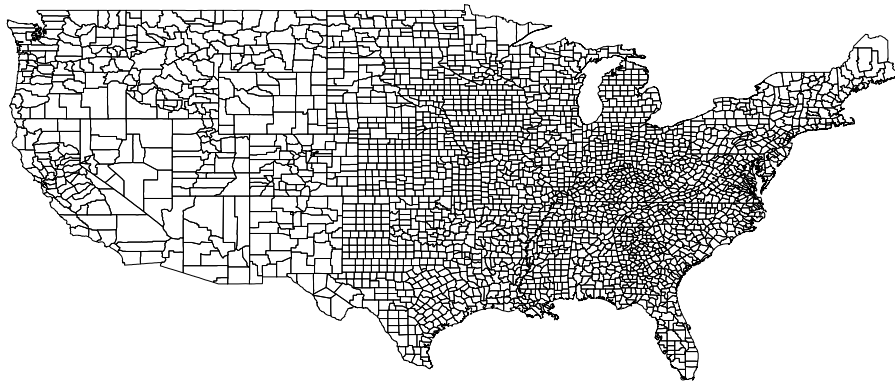
(a) United States
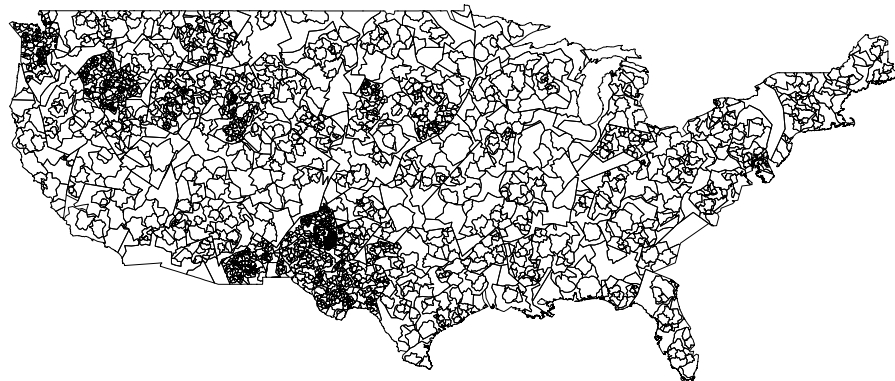
(b) *RI-maps* in United States boundary

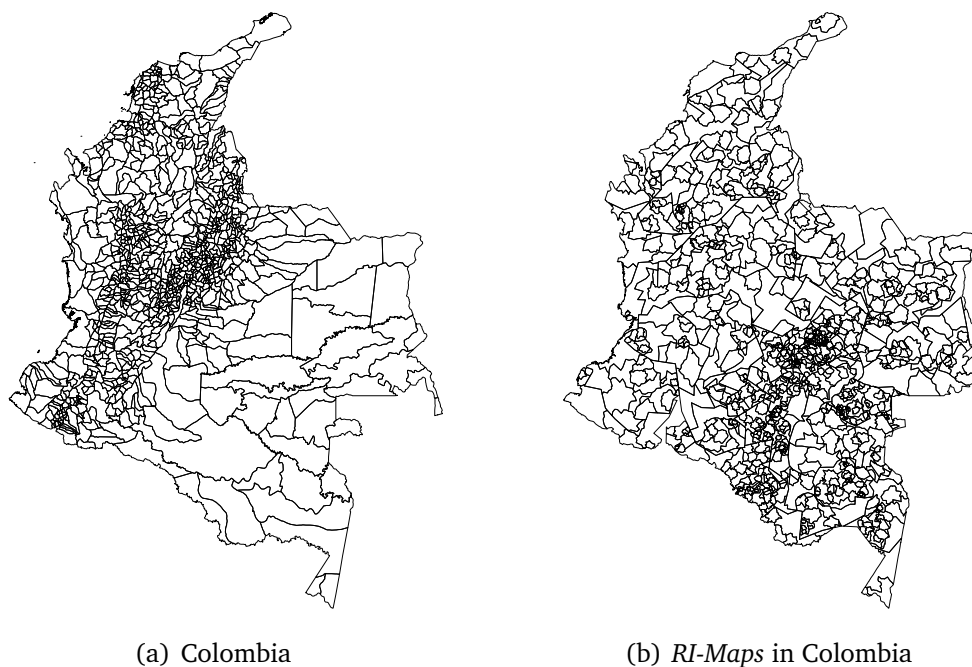**Fig. 5.6:** *RI-Map* **using United States external boundary as initial polygon (3.304 polygons)**

(a) Colombia   (b) *RI-Maps* in Colombia

Fig. 5.7: *RI-Map* **using Colombia external boundary as initial polygon (1.124 polygons)**

5.3 Application of *RI-Maps*

By now it is clear the regional scientists necessity of a new tool to generate tilings with similar topological characteristics of real maps, however in this section an example of use is exposed, proposing an improvement to the computational efficiency analysis made by Duque et al. (2010) of the AMOEBA (A Multidirectional Optimal Ecotope-Based Algorithm). In summary AMOEBA is an algorithm originally proposed by Aldstadt and Getis (2006) to detect spatial clusters and posteriorly computationally improved by Duque et al. (2010), where using a constructive algorithm managed to make it feasible in maps with a high number of areas. To present the results Duque et al. (2010) proposed 3 computational experiments one of them analyzing the running time of AMOEBA while the number of areas of regular tilings increase. Understanding AMOEBA is easy to figure that the computational efficiency is highly affected by the maximum number of neighbors handled during the iterative process, which means that regular lattices are not hardly enough computational tests for AMOEBA. To illustrate this, on one hand a

tiling of 900 areas is constructed and over it 30 Spots processes of 4 clusters are generated, on the other hand, a set of 30 *RI-Maps* with 900 areas are constructed, each of them with a Spots process of 4 clusters. Latter AMOEBA is running over each Spot process and the running time recorded, Fig. 5.10 shows the differences in the execution times. This figure reinforce confirms the fact that using regular tilings to infer the computational efficiency of an algorithm could over-evaluate the real behaviour. It is important to note that our experiment is to show that robust conclusions can be made using *RI-Maps* and not invalidate the results obtained in (Duque et al., 2010).

(a) Italy

(b) *RI-Maps* in Italy boundary

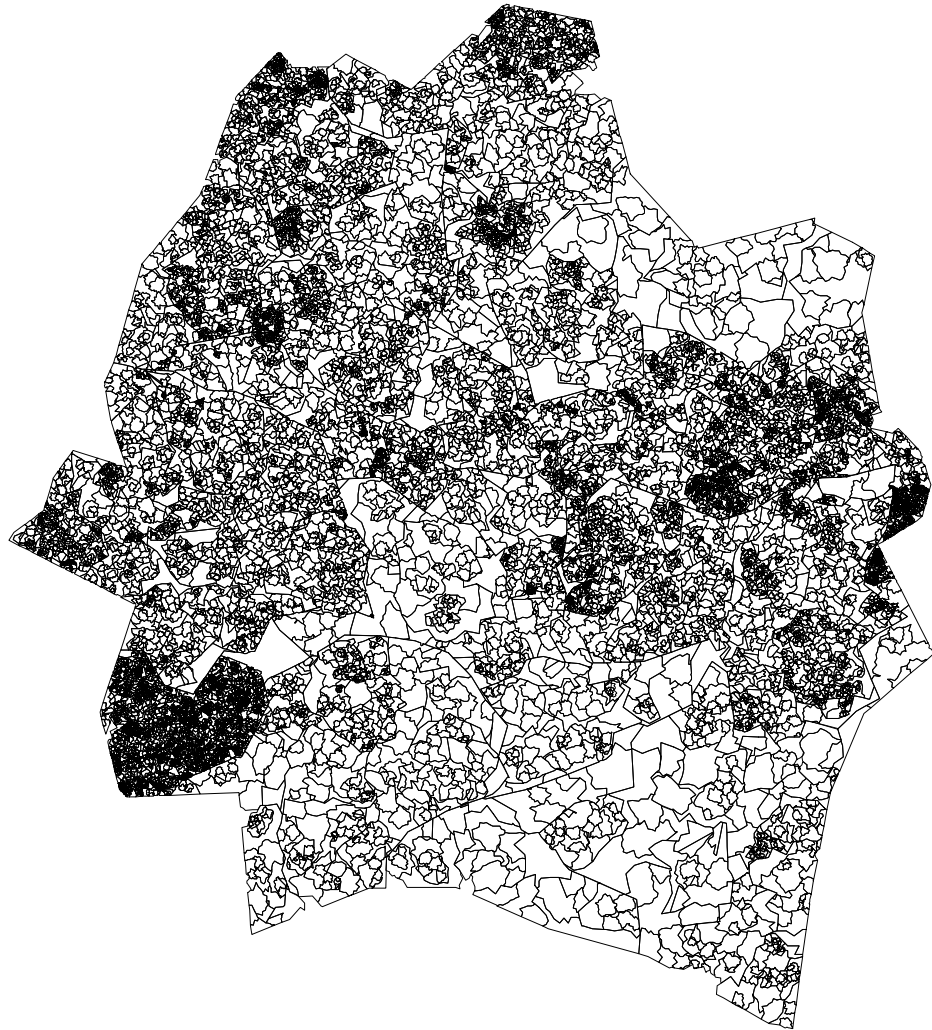**Fig. 5.8:** *RI-Map* **using Italy external boundary as initial polygon (7.308)**

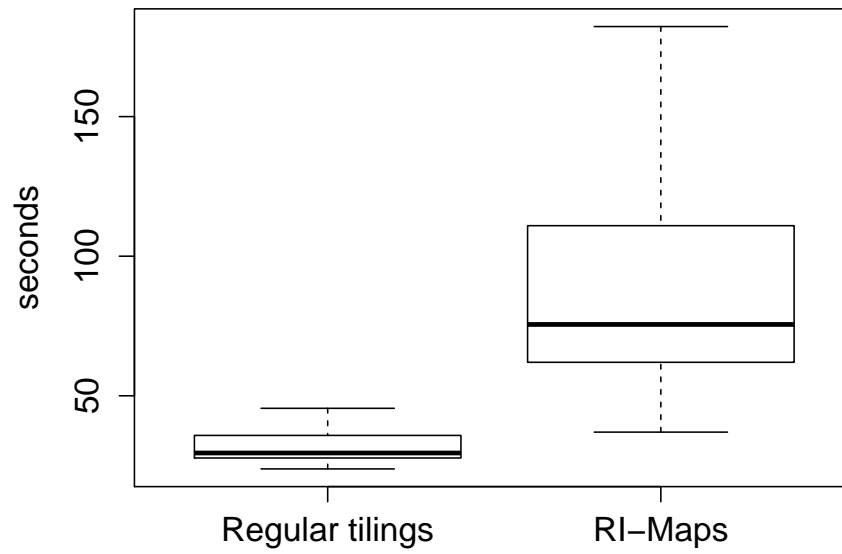**Fig. 5.9:** *RI-Map* **of 20.000 areas**

Fig. 5.10: Execution times of AMOEBA over regular tilings and *RI-Maps* of 900 areas

6. Conclusions

From this project different conclusions can be established, some of them are based on the theoretical background of tilings in regional sciences and others from a computational point of view.

- Based on the topological characteristics of tilings it is evident that commonly used tilings are highly restricted to emulate realistic space partitions. Taking into account these restrictions different consequences of an inappropriate selection of the tilings to test the methodologies emerge, for example mislead the analysis of an indicator or over-evaluate the computational efficiency of an algorithm.

- Despite of the importance of tilings to made a good evaluation of methodologies and algorithms there are not enough tools to allow researchers to properly infer the capabilities of their methods in real applications.

- A new tool to generate realistic tilings based on stochastic processes, fractal theories and geometric computation is proposed in order to allow researches to properly evaluate their algorithms and methodologies.

- From the implementing process in python it was found that currently geo-computational tools are not powerful enough to deal with quite irregular geometries, making difficult to develop these types of algorithms and forcing developers to seek for alternative solutions for simple theoretical geometric operations complicating the codes and making them lose some computational efficiency.

7. Future research

During the development process of *RI-Maps* different challenges and ideas for future researches emerged. Some of them are briefly described bellow.

- In section 4.3.2 the surface of fractal dimension of a mean reverting process is constructed in an empirical way, Fig. 4.4, however an ideal way to get that surface would be through an analytical procedure. The problem of the analytical procedure is that the euclidean distance of an stochastic polygon can not be obtained due to the square root is not a lineal function and is not possible to pass from (7.2) to (7.3), being (7.3) the expression obtained using the stochastic calculus for a mean reverting process. A future research is to use a different distance definition to avoid the square root and state a analytical formulation for the fractal dimension of a mean reverting process (4.13).

$$
\begin{aligned}
d(X_t) &= d(X_0, X_1) + d(X_1, X_2) + d(X_2, X_3) + \cdots + d(X_{T-1}, X_T) & (7.1) \\
&= \sqrt{d^2(X_0, X_1)} + \sqrt{d^2(X_1, X_2)} + \cdots + \sqrt{d^2(X_{T-1}, X_T)} & (7.2) \\
&\neq \sqrt{d^2(X_0, X_1) + d^2(X_1, X_2) + \cdots + d^2(X_{T-1}, X_T)} & (7.3)
\end{aligned}
$$

- As is exposed in section 3.2 the topological characteristics of maps from different parts of the world may differ, see Figure 3.1. Based on these differences a future research is to deeply study how the 7 indicators proposed in section 2.2 change in space.

- Due to the computational problem of geometric packages it is clear the necessity of new improvements to geocomputational methods.

8. Python library

In order to make the algorithms proposed in this thesis available for researchers they are implemented in python taking advantage of the basic functions of clusterpy, library of spatial clustering algorithms developed by RiSE group (Duque et al., 2011b). In this section is explained how to install clusterpy and additional codes required to execute *RI-Maps*, latter a briefly introduction to the basic operations with clusterpy is made and posteriorly some examples are presented.

8.1 Installing clusterpy

ClusterPy is a library of algorithms for spatial clustering, however it has some functions that make easy the use of spatial algorithms, for example, importing and exporting maps. clusterpy (Version 1.1.0) incorporates *MR-Polygons* and *RI-Maps* algorithms. In order to install clusterpy (Version 1.1.0) follow the instructions of *http://rise-group.org/section/Software/clusterPy/*.[12]:

8.2 Examples of clusterpy

Using clusterpy different operations can be made, for example import and export shape files, simulate spatial processes and generate regular tilings, next some examples of use are presented, however for more advanced examples please refer to "http://www.rise-group.org/risem/clusterpy/". Aditionally for more examples of how to create *MR-Polygons* and *MR-Maps* see the section 4.3.3 and 4.4.3 respectively.

Importing shape files Creating a new Layer from a shapefile

```
>>> import clusterpy
>>> layer = clusterpy.importArcData("<SHAPE FILE>")
```

Regular tilings Creating a new Layer with a regular lattice of 10×10

```
>>> import clusterpy
>>> grid = clusterpy.createGrid(10, 10, lowerLeft=(0, 0), upperRight=(100, 100))
```

---

[12]It is assumed that python 2.7 is already installed in your computer and can be executed from the terminal, for more information about python see "http://www.python.org/"

Simulated Spots process Creating a grid of 10×10 with an spots process with 4 clusters, a level of compactness of 0.7, and significance of 0.9

```
>>> import clusterpy
>>> grid = clusterpy.createGrid(10, 10, lowerLeft=(0, 0), upperRight=(100, 100))
>>> grid.generateData("Spots","rook", 1, 4, 0.7, 0.9)
```

Exporting data to Shapefile Creating a grid of 10×10 with an spots process with 4 clusters, a level of compactness 0.7, and significance of 0.9 and exporting it to a shape file

```
>>> import clusterpy
>>> grid = clusterpy.createGrid(10, 10, lowerLeft=(0, 0), upperRight=(100, 100))
>>> grid.generateData("Spots","rook", 1, 4, 0.7, 0.9)
>>> grid.exportArcData("gridFile")
```

Exporting RI-Map to Shapefile Creating a RI-Map of 1.600 and saving it to a shape file

```
>>> import clusterpy
>>> grid = clusterpy.createGrid(10, 10, lowerLeft=(0, 0), upperRight=(100, 100))
>>> grid.generateData("Spots","rook", 1, 4, 0.7, 0.9)
>>> grid.exportArcData("gridFile")
```

BIBLIOGRAPHY

Aldstadt, J. and Getis, A. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis*, 38(4):327–343.

Allen, M., Brown, G. J. G., Miles, N. N. J., and Martin, A. (1995). Measurement of boundary fractal dimensions: review of current techniques. *Powder technology*, 84(1):1–14.

Anselin, L. (1988). *Spatial economtrics: methods and models*. Kluwer Academic Publishers, Dordrecht, 1 edition.

Anselin, L. and Moreno, R. (2003). Properties of tests for spatial error components. *Regional Science and Urban Economics*, 33(5):595–618.

Aste, T., Szeto, K., and Tam, W. (1996). Statistical properties and shell analysis in random cellular structures. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 54(5):5482–5492.

Barnsley, M. F. (2000). *Fractals everywhere*. Morgan Kaufmann.

Bartlett, M. S. (1975). *Probability, statistics, and time: a collection of essays*. Chapman and Hall, New York, 1 edition.

Batty, M. and Longley, P. (1994). *Fractal cities: A geometry of form and function*. Harcout Brace & Company, Publishers, London.

Berger, R. (1966). *The undecidability of the domino problem*. Amer Mathematical Society.

Boots, B. (1982). Comments on the use of eigenfunctions to measure structural properties of geographic networks. *Environment and Planning A*.

Boots, B. (1984). Evaluating principal eigenvalues as measures of network structure. *Geographical Analysis*, 16(3):270–275.

Boots, B. (1985). Size effects in the spatial patterning of nonprincipal eigenvectors of planar networks. *Geographical Analysis*, 17(1):74–81.

Boyce, R. R. and Clark, W. A. V. (2012). The concept of shape in geography. *Society*, 54(4):561–572.

Chen, Y. (2011). Derivation of the functional relations between fractal dimension of and shape indices of urban form. *Computers, Environment and Urban Systems*.

Chiu, A. S. N. (2011). Limit Theorems for the Time of completion of Johnson-Mehl tessellations. *Advances*, 27(4):889–910.

Cole, J. (1964). Study of major and minor civil divisions in political geography. In *20th International Gepgraphical Congress*, Sheffield. University of Nottingham.

Coxeter, H. S. M. (1974). *Regular complex polytopes*. CUP Archive.

Douglas, D. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122.

Duque, J., Aldstadt, J., Velasquez, E., Franco, J., and Betancourt, A. (2010). A computationally efficient method for delineating irregularly shaped spatial clusters. *Journal of Geographical Systems*, pages 1–18.

Duque, J. C., Anselin, L., and Rey, S. J. (2011a). The max-p-regions problem. *Journal of Regional Science*, 52(3):397–419.

Duque, J. C., Dev, B., Betancourt, A., and Franco, J. L. (2011b). clusterPy: Library of spatially constrained clustering algorithms.

Duque, J. C., Ramos, R., and Surinach, J. (2007). Supervised regionalization methods: A survey. *International Regional Science Review*, 30(3):195–220.

Frankhauser, P. (1998). The fractal approach: analysis the spatial, a new tool for of urban agglomerations. *New Methodological Approaches in the Social Sciences*, 10(1):205–240.

Garrison, W. (1964). Factor-analytic study of the connectivity of a transportation network. *Papers in Regional Science*.

Ghyka, M. (2004). *The geometry of art and life*. Kessinger Publishing.

Gibbs, J. (1961). A method for comparing the spatial shapes of urban units. *Urban research methods*, pages 96–106.

Gould, P. (1967). On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers*, 42(42):53–86.

Griffith, D. (1987). Toward a theory of spatial statistics: Another step forward. *Geographical analysis*, 19(1):69–82.

Grunbaum, B. and Shephard, G. C. (2011). *Tilings and patterns*. Dover Publications.

Haggett, P. (1977). *Locational analysis in human geography*. John Wiley & Sons, London.

Hijmans, R., Guarino, L., Jarvis, A., O'Brien, R., Mathur, P., Bussink, C., Cruz, M., Barrantes, I., and Rojas, E. (2011). Diva-GIS.

Hooper, P. and Hewings, G. (1981). Some properties of space-time processes. *Geographical Analysis*, 13(3):203–223.

Horton, R. (1932). Drainage basin characteristics. *Transactions of the American Geophysical Union*, 13.

Johnson, D. L. (2001). *Symmetries*. Springer.

Kindratenko, V. and Treiger, B. (1996). Chemometrical approach to the determination of the fractal dimension (s) of real objects. *Chemometrics and Intelligent Laboratory Systems*, 34:103–108.

Le Caer, G. and Delannay, R. (1993). The administrative divisions of mainland France as 2D random cellular structures. *Journal de Physique I*, 3(8):1777–1800.

Le Caer, G. and Delannay, R. (1995). Topological models of 2D fractal cellular structures. *Journal de Physique I*, 5(11):1417–1429.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*. W.H. Freeman.

Mao, X. (1997). *Stochastic differential equations and applications*. Horwood publishing limited, Chichester, 1 edition.

Miller, V. (1953). A quantitative geomorphic study of drainage basin characteristics in the clinch mountain area Virginia and Tenesse. *ONR geography branch*, PROJECT NR.

Moller, J. (1995). Generation of Johnson-Mehl crystals and comparative analysis of models for random nucleation. *Advances in applied probability*, 27(2):367–383.

Mur Lacambra, J. (1992). Contrastes de autocorrelación espacial: Un estudio de Monte Carlo. *Estadística española*, 34(130):285–308.

Penrose, R. (1974). The rôle of aesthetics in pure and applied mathematical research. *Journal of The Institute of Mathematics and its Aplications*, 10:266–271.

Peshkin, M., Strandburg, K., and Rivier, N. (1991). Entropic predictions for cellular networks. *Physical review letters*, 67(13):1803–1806.

Radin, C. (1993). Symmetry of tilings of the plane. *American mathematical society*, 29(2):213–217.

Richardson, L. (1961). The problem of contiguity: an appendix of statistics of deadly quarrels. *General systems yearbook*, 6(13):139–187.

Robinson, R. and Berkeley, R. M. R. (1971). Undecidability and nonperiodicity for tilings of the plane. *Inventiones mathematicae*, 12(3):177–209.

Stoddart, D. (1965). The shape of atolls. *Marine Geology*, 3(5):269–283.

Tilley, R. (2006). *Crystals and crystal structures*. Wiley, England, 1 edition.

Tinkler, K. (1972). The physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers*, 55(55):17–46.

Weaire, D. and Rivier, N. (2009). Soap, cells and statistics random patterns in two dimensions. *Contemporary Physics*, 50(1):199–239.

Weeitty, A. (1969). *On the form of drainage basins*. Dept. of Geography, Pennsylvania State University, Pennsylvania, 1 edition.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3):434–449.